Nitrate sensing and uptake in diatoms: from molecular evolution to functional characterisation

Anna Santin

Thesis submitted for the degree of
Doctor of Philosophy (PhD)
in Life, Health and Chemical Sciences

October 2022
Nitrate sensing and uptake in diatoms: 
from molecular evolution to functional characterisation

Anna Santin

External Reviewer: Dr. Flora Vincent  
Department of Developmental Biology  
European Molecular Biology Laboratory, EMBL (Heidelberg, DE)

Internal Reviewer: Dr. Gabriele Procaccini  
Department of Integrative Marine Ecology  
Stazione Zoologica Anton Dohrn (Naples, IT)

Director of Studies: Dr. Maria Immacolata Ferrante  
Department of Integrative Marine Ecology  
Stazione Zoologica Anton Dohrn (Naples, IT)

Internal Supervisor: Dr. Alessandra Rogato  
1. Institute of Bioscience and BioResources, IBBR  
   National Research Council, CNR (Naples, IT)  
2. Department of Integrative Marine Ecology  
   Stazione Zoologica Anton Dohrn (Naples, IT)

External Supervisor: Dr. Angela Falciatore  
Institut de Biologie Physico-Chimique, IBPC  
UMR7141 CNRS - Sorbonne Université (Paris, FR)

October 2022
Abstract

Diatoms constitute one of the most important phytoplankton groups in the ocean, thanks to their adaptative capacity to face environmental variations, among which nutrient availability. To cope with fluctuations, diatoms own sophisticated mechanisms, still largely unknown, to sense and transport nutrients from the external environment and to reallocate them inside the cell.

This PhD thesis provides the first characterisation of the Nitrate/Peptide Transporter Family (NPFs) in diatoms. NPFs are putative low-affinity nitrate transporters, known in other organisms to be active at high nitrate concentrations, rarely found in the ocean. Beside nitrate, NPFs have been shown to recognise a remarkably broad range of diverse substrates in organisms where they were characterised, ranging from di- and tripeptides in bacteria, to a wide variety of different molecules in plants, such as phytohormones. However, for diatom NPFs exploration is still at its infancy.

Using a multilevel approach which integrated omics, phylogenetic, structural and expression analyses, we revealed an evolutionary divergence into two distinct branches, with a different predicted subcellular localisation suggesting functional diversification.

In order to understand the function of diatom NPFs, and to explain the apparent contradiction of the presence of low-affinity nitrate transporters in an environment in which nitrate levels are never very high, we used reverse genetics approaches. We generated overexpressing strains and CRISPR/Cas9 loss of function mutants in the model species *Phaeodactylum tricornutum*.

Functional characterisation of the mutants suggested that the two different *P. tricornutum* NPFs could be respectively required for the regulation of intracellular nitrogen fluxes, especially nitrate reallocation from the vacuole, and for internal pH regulation and ion transport across chloroplast membranes. So, diatom NPFs evolved to regulate intracellular nutrient and ion transport, rather than uptake them from the external environment, adding new pieces to the complex puzzle of diatom physiology which contributes to their ecological success.
Acknowledgements

I would like to thank the Open University (OU) and Stazione Zoologica Anton Dohrn for giving me the opportunity and fellowship to pursue my PhD.

This thesis would never be written without the help of my research supervisors: I want to warmly thank Dr. Maria Immacolata Ferrante, my director of studies, and Dr. Alessandra Rogato, my internal supervisor, for the support, the teaching, the guidance and the encouragement given to me during my PhD.

I would also like to extend my thanks to Dr. Monia Teresa Russo, Dr. Maurizio Ribera d’Alcalà, Dr. Maurizio Chiurazzi and Dr. Angela Sardo for their valuable and constructive suggestions during the planning and the development of this research, for sharing with me their knowledge and experience. I would like to express my gratitude to the post-docs of Ferrante Lab, Pina, Francesco, Antonella and Svenja, for their constant assistance, their advice and their answers. They showed me what a working group is, not only from the professional point of view but also from the human one. I would also thank my wingman Flavio and all the other colleagues of Stazione Zoologica for their constant support.

I would also thank Dr. Angela Falciatore and all the people in her laboratory, at the Institut de Biologie Physico-Chimique in Paris, for hosting me one month and creating such an inspiring and stimulating environment. A special thanks to Dany and Benjamin, for all the time they dedicated to me and my work, and also to Erik and Elena, for being such friendly with me. I also thank the EMBO Fellowship Programme that gave me the opportunity to spend one month there.

A huge thank to all my housemates, who lived this adventure with me every day, and in particular to the crazy Veneto group (Emma, Valentina and Annalisa) for sharing with me smiles, tears, fun, anger, brilliant and bad ideas. I could not explain how grateful I am for time we spent together.

And thanks to my parents and my sister, for teaching me that passion and enthusiasm are essential components of everyday life, I wish to make them always proud of me. Finally, I want to thank Michele, for supporting and standing by me through all this incredible adventure, I would not be here without him.
Summary

Chapter 1: Introduction

1.1. Diatoms

1.1.1. Life in the Ocean

1.1.2. Phytoplankton

1.1.3. Diatom place in the world

1.1.4. Diatom morphology and structure

1.1.5. Diatom evolution

1.1.6. Diatoms as result of secondary endosymbiosis

1.1.7. Diatom cell cycle

1.1.8. Ecological contributions of diatoms

1.1.9. Economic and biotechnological importance of diatoms

1.2. Molecular resources

1.2.1. Omics resources

1.2.2. Meta-omics revolution

1.2.3. Genetic tools for diatoms

1.2.4. Genome editing

1.2.5. The model diatom Phaeodactylum tricornutum

1.3. Nutrient transporters

1.3.1. To cope with a variable environment

1.3.2. Kinetics of nutrient transporters

1.3.3. Nutrients and their specific transport

1.3.4. Nitrogen transporters in diatoms

1.3.5. Ion balance in diatoms

1.3.6. Nutrient sensing

1.4. Aim of the thesis

Chapter 2: Diatom NPFs identification, evolution and structure

Abstract

2.1. Introduction

2.2. Materials and Methods

2.2.1. diNPFs identification in public genomes, TARA Oceans and MMETPS databases

2.2.2. diNPFs alignment
2.2.3. Phylogenetic analyses ................................................................. 64
2.2.4. Global meta-omics of MATOU diNPFs ........................................... 65
2.2.5. Culture conditions ..................................................................... 65
2.2.6. RNA extraction, primer design and qPCR ....................................... 65
2.2.7. In silico analysis of the diNPFs non-coding sequences ....................... 67
2.2.8. Structural prediction of diNPFs ........................................................ 67
2.2.9. Prediction of subcellular localisation ................................................. 68
2.3. Results ............................................................................................ 68
   2.3.1. diNPF genes in diatom genome, TARA Oceans and MMETPS databases 68
   2.3.2. Phylogeny of diNPFs ................................................................. 73
   2.3.3. diNPFs distribution in the global ocean ........................................... 75
   2.3.4. diNPFs expression data based on literature ...................................... 80
   2.3.5. Expression patterns of NPFs in P. tricornutum ................................ 85
   2.3.6. Upstream regulation of diNPF genes ............................................ 86
   2.3.7. Structural modelling of diNPFs ...................................................... 89
   2.3.8. Predicted subcellular localisation ................................................. 96
2.4. Discussion ....................................................................................... 98

Chapter 3: Proteolistic protocol optimisation for Ptnpfs knock-out mutants generation ............................................................. 105

Abstract ........................................................................................................ 106
3.1. Introduction ....................................................................................... 107
3.2. Materials and Methods ................................................................. 110
   3.2.1. CRISPR-Cas9 design and constructs .............................................. 110
   3.2.2. crRNA::tracrRNA duplexes forming gRNAs .................................... 111
   3.2.3. Diatom culture and plating ......................................................... 112
   3.2.4. Assembly of RNP complexes ...................................................... 113
   3.2.5. Cas9 and gRNAs in vitro functional validation ................................ 114
   3.2.6. Preparation of gold nanoparticles and loading of RNP complexes onto them... 115
   3.2.7. Transformation and replating ..................................................... 116
   3.2.8. PCR analysis on transformed knock-out cells and Sanger sequencing ...... 117
3.3. Results ............................................................................................. 118
   3.3.1. Drug selection set up .................................................................. 118
   3.3.2. gRNAs design and in vitro Cas9 assays ...................................... 119
   3.3.3. Pipeline for proteolistic mutants generation and screening .............. 124
3.3.4. *Ptapt* knock-out mutants generation and screening ................................................. 127
3.3.5. *Ptapt-Ptnpf1* knock-out mutants generation and screening ................................. 130
3.3.6. *Ptapt-Ptnpf2* knock-out mutants generation and screening ............................... 135
3.3.7. *Ptapt-Ptnpf1-Ptnpf2* double knock-out mutants generation and screening .......... 139
3.3.8. Consecutive proteolistics to generate *Ptapt-Ptnpf1-Ptnpf2* double knock-out mutants ........................................................................................................................................... 141
3.4. Discussion ....................................................................................................................... 142

Chapter 4:  *P. tricornutum* NPF1 functional characterisation ................................. 149

Abstract .................................................................................................................................. 150
4.1 Introduction .......................................................................................................................... 151
4.2. Materials and Methods ..................................................................................................... 154
4.2.1. Diatom cultures........................................................................................................ 154
4.2.2. *In silico* prediction of PtNPF1 vacuolar targeting ................................................... 154
4.2.3. PtNPF1 overexpression vectors .............................................................................. 154
4.2.4. Biolistic transformation of *P. tricornutum* .............................................................. 156
4.2.5. PCR screening on transformed cells overexpressing PtNPF1 ................................. 157
4.2.6. Confocal Microscopy .............................................................................................. 158
4.2.7. Vacuole morphology measurements......................................................................... 158
4.2.8. Growth curves and N starvation and repletion experiment .................................... 159
4.2.9. Extra- and intra-cellular NO\textsubscript{3}\textsuperscript{-} content analyses......................... 160
4.2.10. Statistical analyses ................................................................................................. 161
4.2.11. RNA extraction and qPCR analysis....................................................................... 161
4.2.12. Vector for PtNPF1 heterologous expression in *Xenopus laevis* ............................ 163
4.3. Results ................................................................................................................................ 164
4.3.1. A di-leucine-based motif important for vacuolar targeting found in PtNPF1 ...... 164
4.3.2. Generation and selection of *P. tricornutum* PtNPF1 overexpressing strains ........ 165
4.3.3. PtNPF1 subcellular localisation................................................................................. 169
4.3.4. Vacuole behaviour in N starvation and different N sources ................................. 172
4.3.5. Ptnpf1 knock-out phenotype in response to N repletion ........................................ 176
4.3.6. Gene expression analyses ...................................................................................... 181
4.4. Discussion ....................................................................................................................... 184

Chapter 5:  *P. tricornutum* NPF2 functional characterisation ................................. 191

Abstract .................................................................................................................................. 192
5.1. Introduction ....................................................................................................................... 193
5.2. Materials and Methods......................................................................................................................... 197
5.2.1. Diatom cultures........................................................................................................................................ 198
5.2.2. PtNPF2 overexpression vectors........................................................................................................... 198
5.2.3. PCR analysis on transformed cells overexpressing PtNPF2 ............................................................... 200
5.2.4. RNA extraction and qPCR analysis....................................................................................................... 201
5.2.5. Confocal Microscopy............................................................................................................................ 202
5.2.6. Growth curves in different conditions................................................................................................. 202
5.2.7. Normal to low pH shift ....................................................................................................................... 203
5.2.8. Dark to light shift ................................................................................................................................. 204
5.2.9. Photosynthetic measurements........................................................................................................... 204
5.2.10. Mini-FIRE measurements ............................................................................................................... 205
5.2.11. Intracellular pH measurements ....................................................................................................... 206
5.2.12. Bicarbonate assay ........................................................................................................................... 206
5.2.13. Statistical analyses............................................................................................................................. 207
5.2.14. Sample preparation for RNA-seq ...................................................................................................... 207
5.2.15. Vector for PtNPF2 heterologous expression in Xenopus laevis ......................................................... 207

5.3. Results......................................................................................................................................................... 208
5.3.1. Generation and selection of P. tricornutum PtNPF2 overexpressing strains ........................................ 208
5.3.2. In vivo subcellular localisation of PtNPF2 ........................................................................................ 212
5.3.3. Phenotypic analyses of mutants ........................................................................................................ 215
5.3.4. Normal to low pH shift ....................................................................................................................... 222
5.3.5. Dark to light shift ................................................................................................................................. 232
5.3.6. Bicarbonate assay ............................................................................................................................... 234
5.3.7. Samples for RNA-seq ........................................................................................................................ 235

5.4. Discussion................................................................................................................................................ 237

Chapter 6: General discussion ..................................................................................................................... 245
6.1. Thesis aim and main results .................................................................................................................... 246
6.2. Thesis discussion ...................................................................................................................................... 248
6.3. Future perspectives .................................................................................................................................. 256

References ...................................................................................................................................................... 263

Appendices ..................................................................................................................................................... 289
Appendix A. Supplementary materials relative to Chapter 2 ..................................................................... 290
Supplementary Table S2.1 ............................................................................................................................ 290
Supplementary Table S2.2 ............................................................................................................................ 298
List of Figures

Figure 1.1. Major evolutionary and biogeochemical events during the history of life on Earth, since the evolution of oxygenic photosynthesis, and how they affected atmospheric O$_2$ trends. Light blue lines represent glaciation events, while grey ones represent the estimated span of the events cited. To note that because the oldest direct measurements of atmospheric O$_2$ come from Pleistocene ice cores, the detail in the Phanerozoic curve is based on models. Image adapted from Benoiston et al. (2017). ....................................... 3

Figure 1.2. Schematic representation of diatom morphology. On the left, representation of frustule morphological features. On the right, organelles organisation inside diatom cell. .................................................................................................................................... 6

Figure 1.3. Estimated timing of divergence of the four major diatom lineages and coincident events in Earth history. Maps are palaeographic reconstructions of continent locations during the emergence of the diatom lineages. Shallower depths in the ocean are indicated by lighter blues. Adapted from Armbrust (2009). ............................................. 8

Figure 1.4. Schematic representation of diatom evolution. The “melting pot” diatom genome finds its origins in successive gene transfers following endosymbiosis between red and green algae and a host heterotrophic cell (Moustafa et al., 2009; Benoiston et al., 2017). The event is proposed to have permitted gene transfer from prey nuclei to host nucleus as well as from organelles to nucleus. Diatoms also seem to have acquired genes through lateral gene transfer both before and after the diversification of pennates and centrics. Image adapted from Bowler et al. (2010). ........................................................ 10

Figure 1.5. Schematic drawing of the life cycle of a centric and a pennate diatom. Diatom cells are diploid and are surrounded by a rigid frustule made of two unequal thecae. During mitosis, the new thecae are formed inside the frustule, causing a progressive decrease in the population cell size. Once a species-specific size threshold is reached, the formation of gametes takes place. Conjugation of the haploid gametes produces a zygote that expands into an auxospore, from which the large initial cell is synthesised.........13

Figure 1.6. Extent of diatom-rich sediments compared with the distribution of modern diatoms in the ocean. Biosiliceous sediments are present in regions that, still today, are largely dominated by diatoms, in particular the Southern Ocean. A) Small dots represent seafloor sediment samples defined as containing predominantly diatom sediments, muds, also mixed with calcareous one. These data come from the Index to Marine and Lacustrine Geological Samples (Curators of Marine and Lacustrine Geological Samples Consortium, 2014; Dutkiewicz et al., 2015b). Circles of varying size and blue color correspond to diatom relative abundances determined by the Marine Ecosystem Model Inter-Comparison Project (Leblanc et al., 2012) and by the TARA Oceans survey (Malviya et al., 2016), based on metabarcoding data coming from the photic zone (both subsurface SRF and deep chlorophyll maximum DCM) and corresponding to four size classes (0.8–5 μm, 5–20 μm, 20–180 μm, and 180–2 000 μm). B) Water column inventory of diatom biomass (mmol C/m$^2$) from a biogeochemical/ecosystem simulation (Dutkiewicz et al., 2015a). Image adapted from Benoiston et al. (2017).................................................................. 15

Figure 1.7. Scheme of the different uses of diatoms for green industry, from the application of frustule optical properties to the environmental technologies and the different uses of metabolites produced by diatoms. Image inspired to Sharma et al. (2021b). ......................................................................................... 18
Figure 1.8. The map shows the cruise track of TARA Oceans as she sailed the world from September 2009 to December 2013 and the location of 210 stations, which were chosen to cross and to sample as many biogeographic provinces and environmental features as possible (sea surface temperature shown as a colour gradient). Overall, more than 35 000 samples of seawater and plankton were collected and archived in partner laboratories. Image adapted from Sunagawa et al. (2020).

Figure 1.9. Most applied transformation methods in diatoms: A) microparticles bombardment, also called biolistic method; B) electroporation, which exploits electric pulses to create transient pores in the cell membrane; and C) bacterial conjugation, through the episome transfer the bacteria to the diatom cell.

Figure 1.10. Schematic representation of the CRISPR/Cas9 system.

Figure 1.11. The pennate diatom P. tricornutum showed in its three main morphotypes through light microscope: left, fusiform; top right, triradiate; bottom right, oval. Image adapted from Vardi et al. (2009).

Figure 1.12. Timeline of the major publications and discoveries dealing with the model diatom P. tricornutum.

Figure 1.13. Average spatial concentration of nitrates, silicates and phosphate at surface and at 200 meters depth. Scale colours are not homogeneous. Data are taken from the World Ocean Atlas of Climatology (https://www.ncei.noaa.gov/access/world-ocean-atlas-2018f/) and refer to year 2018.

Figure 1.14. Major classes and abundance of macronutrient transporters from diatom genomes. Genes were identified from the literature and transporter annotations from the Transporter Classification Database (TCDB) obtained from Phycocosm (a comparative algal genomics resource) for ten diatoms (Smith and Allen, 2022). Box and whisker plot shows distribution of data into quartiles, with the central line indicating the exclusive median, and x indicates the mean. Data labels show the mean value. Image adapted from Smith and Allen (2022).

Figure 1.15. Compartmentalised model of diatom nutrient transport and storage. Colour gradients indicate putative dissolved organic matter, nutrient, or pH gradients. See text for abbreviations. NO$_3^-$: nitrate, NO$_2^-$: nitrite, NH$_4^+$: ammonium, P$_i$: inorganic phosphate, Si(OH)$_4$: silicic acid, CO$_2$: carbon dioxide, HCO$_3^-$: bicarbonate, Fe$^{2+}$ and Fe$^{3+}$: ferrous and ferric states of iron, Fe$^{3+}$-sid: iron-bounded siderophore, HPO$_4^{2-}$: hydrogen phosphate. NRT2: high-affinity Nitrate Transporter, NPF: low-affinity Nitrate Transporter, AMT: Ammonium Transporter, UT: Urea Transporter, AAT: Amino Acids Transporter, HP$_i$ and NaP$_i$: H$^+$/P$_i$ and Na$^+$/P$_i$ Transporters, SIT: Silicon Transporter, LSI2: Low Silicon Rice 2, SLC: Solute Carrier, FBP/FRE: Ferrochrome Binding Protein/Ferric Reductase, ISIP: Iron Starvation-Induced Protein, CLC: Chloride Channel, VPT: Vacuolar Phosphate Transporter, SDV: Silica Deposition Vesicle. Image created with BioRender, following Smith and Allen (2022) and Brownlee et al. (2022).

Figure 1.16. N and the diatom cell. Uptake from the extracellular environment occurs via different transporters, N internalised is then reduced through N assimilation pathways. NO$_3^-$: nitrate, NO$_2^-$: nitrite, NH$_4^+$: ammonium, Glu: glutamate, Gln: Glutamine. NRT/NPF: Nitrate Transporters, NR: Nitrate Reductase, NAR1: Nitrite Transporter, NiR: Nitrite Reductase, GS: Glutamine Synthetase, GOGAT: Glutamine Oxoglutarate Amino Transferase.
Figure 2.1. NPFs evolutionary relationships inferred using the maximum-likelihood and Bayesian inference approaches. Numbers over the nodes represent bootstrap values. For all bootstrap values greater than 75, a posterior probability greater than 0.75 was also found. Branches were collapsed at high taxonomical levels. Salmon box indicates the NPFs Clade I, while cyan box indicates the NPFs Clade II. Red branches are used for diNPFs Clades I and II, yellow for Metazoa, green for plants and green algae, grey for bacteria. H. sapiens: Homo sapiens, P. danica: Pteridomonas danica, A. anophagefferens: Aureococcus anophagefferens, C. wailesii: Coscinodiscus wailesii, C. elegans: Chaeonabditis elegans, P. patens: Physcomitrella patens, R. marina: Rhizochromulina marina. Complete list of sequences is reported in Supplementary Table S2.1. Image from Santin et al. (2021b). ........................................................................ 75

Figure 2.2. diNPF richness (A), expressed as the number of different sequences retrieved in different sampling stations, and diNPF DNA (B) and mRNA (C) relative abundance in the global ocean. Data mapped are from the TARA Oceans dataset, and come from two size classes, 5–20 μm and 20–180 μm (Busseni et al., 2019), and for two different sampling depths: surface (SUR) around 3–5 m depth and Deep Chlorophyll Maximum (DCM) tens of meters below surface. Sampling stations are coloured according to the belonging biome. While in panel A the number of sequences retrieved is presented for all size fractions and depths, in panels B and C, data are separately represented. White dots indicate TARA Oceans stations where no diNPFs were present. Circle size is proportional to abundances. Image from Santin et al. (2021b)............................................................ 78

Figure 2.3. Correlation by means of Pearson’s r of diNPF DNA (Metagenomics, A and B) and mRNA (Metatranscriptomics, C and D) with NO₂⁻ levels in the global ocean for the size class 20–180 μm, at both DCM and SUR sampling depths. In both cases, DNA and mRNA levels have been correlated one to the other and in relation to NO₂⁻ level (E and F). Red circles in the mRNA versus NO₂⁻ scatter plots indicate Type I and Type II responses. Image from Santin et al. (2021b)................................................................... 79

Figure 2.4. Correlation by means of Pearson’s r of diNPF DNA (Metagenomics, A, B and C) and mRNA (Metatranscriptomics, D, E and F) with sunshine duration (SSD) levels in the global ocean for the size class 20–180 μm, at both DCM and SUR sampling depths. In both cases, DNA and mRNA levels have been correlated one to the other and in relation to SSD level (G, H and I). Red circles in the mRNA versus SSD scatter plots indicate Type I and Type II responses. Image from Santin et al. (2021b).................................... 80

Figure 2.5. Expression profile analysis of PtNPFs genes as assessed by qPCR in A) different N concentrations, in different N sources, in light or dark condition, at different pHs, and B) in different ecotypes (Pt1 as control and Pt4). Experiments were performed separately, using as control P. tricornutum wild-type Pt1, grown in 882 μM NO₃⁻ at pH 8, collected during the light phase. Blue bars: logFC (fold-changes) for PtNPF1; red bars: logFC for PtNPF2. Black bars represent standard deviations. Light-blue dotted lines indicate significance thresholds. Image adapted from Santin et al. (2021b). ............................. 88

Figure 2.6. Potential cis-regulatory motifs in the diNPF 5’-flanking regions with their sequence localisation, for P. tricornutum, P. multistriata, T. pseudonana and F. cylindrus. The discovered motifs are indicated on each 5’-flanking sequence of diNPF genes, highlighted in different bar colours and with a bar height equal to their significance. The sequences have different lengths because they correspond to the entire 5’-intergenic region flanking the diNPF genes. Image from Santin et al. (2021b). ............... 86

xiii
Figure 2.7. Separated and overlapped structural models of *P. tricornutum* PtNPF1 (Clade II) and PtNPF2 (Clade I). A) The PtNPF1 model (N-terminal in blue, C-terminal in magenta and lateral helix in yellow) was obtained using the crystal structure of the *A. thaliana* NPF6.3 (pdb: 4oh3) as template (Sun et al., 2014). B) The PtNPF2 model was obtained using both the AtNPF6.3 and the *S. oneidensis* peptide transporter PepTso (pdb: 4uvm) (Fowler et al., 2015) as template structures. C) The two structural models overlapped to highlight their structural differences between diNPF belonging to two different clades. Image from Santin et al. (2021b).

Figure 2.8. Zoomed in view of the putative NO$_3^-$ binding site of PtNPF1. Residues Glu194, Glu197 and Arg198 belong to the conserved ExxER motif. Residues Tyr202 and Tyr439 form part of the substrate binding site. TMHs belonging to the N- and C-terminal domains are in light blue (left) or light grey (right), respectively. Image from Santin et al. (2021b).

Figure 2.9. Transmembrane topology of bacterial, plant and diatom POTs/NPFs. Topology plots showing the 12 transmembrane portions of A) *Shewanella oneidensis* PepT1, based on its crystal structure (pdb: 4uvm) (Fowler et al., 2015); B) *P. tricornutum* NPF2, based on its structural model; C) *Arabidopsis thaliana* NPT6.3, based on its crystal structure (pdb: 4oh3) (Sun et al., 2014); D) *P. tricornutum* NPF1, based on its structural model. ExxER motif amino acids are indicated by red squares; arginine/lysine in blue; glutamic acid/aspartic acid in red. Residues forming a salt bridge between TMH1 and TMH7 or TMH4 and TMH10 are enclosed within a yellow or green circle, respectively. Figure created with the Protter web application (http://wlab.ethz.ch/protter). Image from Santin et al. (2021b).

Figure 2.10. Proposed model for diNPFs functioning, with alternating-access mechanism. Dashed red lines represent salt bridges: the TMH4–TMH10 salt bridge is present in both diNPFs clades, while TMH1–TMH7 salt bridge is only present in Clade I diNPFs. Small yellow circles represent protons, while large blue circles represent a substrate, which could be NO$_3^-$ or other molecules. Image from Santin et al. (2021b).

Figure 3.1. The mechanisms of CRISPR/Cas9-mediated genome editing and DSBs repair. CRISPR/Cas9 can introduce DSBs in DNA. DSBs are repaired by either NHEJ or HDR. Insertions, deletions or other alterations of DNA will occur during this process to achieve gene modifications. Image created with BioRender.

Figure 3.2. Improved proteolistic protocol workflow. Main steps are indicated, with the approximate time required. Created with BioRender. Image from Russo et al. (2022).

Figure 3.3. *P. tricornutum* *Ptl* wild-type and *Ptapt* exposed to different 2-FA concentrations. Cells plated on solid F/2 medium without silica, with adenine (ADE) and with A) 50 µM, B) 30 µM, C) 20 µM, D) 10 µM and E) no 2-FA; F) cells growing in liquid F/2 medium without silica, with adenine and with 20 µM, 10 µM and no 2-FA.

Figure 3.4. *in vitro* Cas9 assay with *PtAPT* gRNAs: gAPT_1 cuts *PtAPT* amplicon into two fragments of 455 bp and 105 bp, while gAPT_3 cuts it into two fragments of 225 bp and 335 bp. M represents the 100 bp marker.

Figure 3.5. gRNAs designed and correspondence to transmembrane helices in the folded protein. A) yellow bars represent the gRNAs located on the *PtNPF1* gene sequence with
bold underlined PAM sequence, B) predicted PtNPF1 protein structure, with residues included between the two designed gRNAs in red. ...................................................................................... 121

Figure 3.6. *in vitro* Cas9 assay with *PtNPF1* gRNAs. A) *PtNPF1* cut only with gNPF1_a (lane 3) and gNPF1_b (lane 4); B) effect of RNP complexes made combining both *PtAPT* (lane 3) and *PtNPF1* (lane 4) gRNAs, and their cut on *PtAPT* and *PtNPF1* fragments. M represents the 1Kb marker. ........................................................... 122

Figure 3.7. gRNAs designed and correspondence to TMH4 and TMH5 in the folded protein. A) yellow bars indicate the gRNAs located on the *PtNPF2* gene sequence with bold underlined PAM sequence, B) predicted PtNPF2 protein structure, with residues included between the two designed gRNAs in red. ...................................................................................... 123

Figure 3.8. *in vitro* Cas9 assay with *PtNPF2* gRNAs: gNPF2_a cuts *PtNPF2* into two fragments of 1456 bp and 405 bp, while gNPF2_b cuts *PtNPF2* into two fragments of 1127 bp and 734 bp. M represents the 1Kb marker. ..................................................... 123

Figure 3.9. Pipeline for proteolistic mutants screening. A) Collection, lysis and gDNA extraction of clones positive to selection, B) PCR screening and C) visualisation of amplicons at different height on agarose gel, D) eventual cloning of single amplicons, E) Sanger sequencing, F) visualisation of chromatograms and G) alignment with wild-type reference gene, H) eventual SNPs study, I) prediction of resulted mutated protein. All these steps are described in detail in the text below...................................................... 124

Figure 3.10. Example of proteolistic mutants screening by PCR. Different amplicon patterns can be observed on the resulting agarose gel: they indicate that different hypothetical mutations occurred, which need to be subsequently tested through sequencing. M represents the marker................................................................. 126

Figure 3.11. Selection and identification of *Ptapt* proteolistic mutants. A) Positive C+ and negative C- control plates with non-transformed wild-type cells growing on F/2 medium without and with selection respectively, and two plates with transformed cells growing on selection. B) agarose gel showing the PCR screening on transformed cells positive to selection: M represents the 100 bp marker and C- the blank. C) *Ptapt* clone 4 chromatogram showing deletion: Cas9 cut genomic DNA in correspondence of the two designed gRNAs (underlined, with PAM sequence in the black squares), the green and pink shadows represent the upstream and downstream regions where wild-type and mutant strains correctly align. ....................................................................................... 128

Figure 3.12. *Ptapt* clone 4 mutation resulting in a truncated protein. A) Translated mutated sequence aligned with wild-type reference amino acid sequence, showing the translation halt; B) predicted protein structures of wild-type and *Ptapt* clone 4 APT proteins. ................................................................................................................................... 129

Figure 3.13. *Ptapt* clone 1 subclones further screened by PCR, still highlight three amplicons on agarose gel. M represents the 100 bp marker and C- the blank. ................................................................. 130

Figure 3.14. Selection and first PCR screening of *PtNPF1* proteolistic mutants. A) Positive C+ and negative C- control plates with not transformed wild-type cells growing on F/2 medium without and with selection respectively, and four plates with transformed cells growing on selection. B) agarose gels showing the results of the PCR screening on *PtAPT* and *PtNPF1* genes on wild-type and transformed cells positive to selection, M represents the 100 bp marker (above and the 1Kb marker (below), while C- represents the blank. ................................................................. 131
Figure 3.15. A) PtNPF1 gene on *P. tricornutum* wild-type genome showing SNPs indicated by red arrows. B) Comparison between wild-type chromatogram showing SNPs and *Ptnpf1* mutant chromatogram which does not present SNPs up- and downstream of the mutation.

Figure 3.16. *P. tricornutum* wild-type and *Ptnpf1* mutants schemes and chromatograms showing Cas9 different cuts on the two *Ptnpf1* alleles. A) PtNPF1 wild-type sequence showing gRNAs and primer couples used for sequencing. B) *Ptnpf1* 2.8 mutation scheme and comparison between wild-type and *Ptnpf1* 2.8 chromatograms showing a 214 bp insertion on allele 1 and a 403 bp deletion on allele 2. C) *Ptnpf1* 2.9 mutation scheme and comparison between wild-type and *Ptnpf1* 2.9 chromatograms showing a 403 bp insertion on allele 1 and a 650 bp deletion on allele 2. gNPF1_a is underlined, including PAM sequence in the black box.

Figure 3.17. Structural models of *P. tricornutum* NPF1 wild-type and knock-out. A) Structural model of PtNPF1 wild-type. B) Truncated protein model of PtNPF1 in the *Ptnpf1* 2.8 strain resulting from the two mutations on the two alleles. C) Truncated protein models of PtNPF1 in *Ptnpf1* 2.9 strain resulting from the two mutations on the two alleles.

Figure 3.18. Selection and first PCR screening of *Ptnpf2* proteolistic mutants. A) Positive C+ and negative C- control plates with non-transformed wild-type cells growing on F/2 medium without and with selection respectively, and four plates with transformed cells growing on selection. B) agarose gels showing the results of the PCR screening on PtAPT and PtNPF2 genes on wild-type and transformed cells positive to selection, M represents the 100 bp marker (above and the 1Kb marker (below), while C- represents the blank.

Figure 3.19. *P. tricornutum* wild-type and *Ptnpf2* mutants schemes and chromatograms showing Cas9 different cuts on the two *Ptnpf2* alleles. A) PtNPF2 wild-type sequence showing gRNAs and primer couples used for sequencing. B) *ptnpf2* 1.15 mutation scheme and comparison between wild-type and *Ptnpf2* 1.15 chromatograms showing a 231 bp insertion on allele 1 and a 334 bp deletion on allele 2. C) *Ptnpf2* 1.16 mutation scheme and comparison between wild-type and *Ptnpf2* 1.16 chromatograms showing a 5 bp insertion and an “A” insertion on allele 1 and a 1181 bp deletion on allele 2. gNPF2_a is underlined, including PAM sequence in the black box.

Figure 3.20. Structural models of *P. tricornutum* NPF2 wild-type and knock-out. A) Structural model of PtNPF1 wild-type. B) Truncated protein models of PtNPF2 in the *Ptnpf2* 1.15 strain resulting from the two mutations on the two alleles. C) Truncated protein model of PtNPF2 in *Ptnpf2* 1.16 strain resulted from the two mutations on the two alleles.

Figure 3.21. Selection and first PCR screening of Ptapt-*Ptnpf1*-*Ptnpf2* proteolistic mutants (also called *Ptnpf1-2* knock-out mutants). A) Positive C+ and negative C- control plates with non-transformed wild-type cells growing on F/2 medium without and with selection respectively, and four plates with transformed cells growing on selection. B) agarose gels showing the results of the PCR screening on PtNPF1 and PtNPF2 genes on wild-type and transformed cells positive to selection, M represents the 1Kb marker and C- the two blanks.

Figure 3.22. *P. tricornutum* wild-type and *Ptnpf1-2* mutants chromatograms showing Cas9 different cut. Comparison between wild-type and *Ptnpf1-2* knock-out strain 1.29 chromatograms showing A) a 403 bp deletion on only one *PtNPF1* allele and B) a 403 bp deletion on only one *PtNPF1* allele and B) a 403 bp deletion on only one *PtNPF1* allele and B) a 403 bp deletion on only one *PtNPF1* allele.
bp insertion on only one PtNPF2 allele. Comparison between wild-type and Ptnpf1-2 knock-out strain 1.30 chromatograms showing C) 229 bp and 1 bp insertions on PtNPF1 alleles and D) a 581 bp deletion on only one PtNPF2 allele. gRNAs are underlined, including PAM sequence in the black boxes.

Figure 3.23. P. tricornutum Pt1 wild-type and Ptnpf1-2 exposed to different 5-FOA concentrations. Cells plated on solid F/2 medium without silica, with uracil (URA) and with A) 100 µg/ml, B) 300 µg/ml, C) 450 µg/ml, D) 600 µg/ml, E) 750 µg/ml and F) 900 µg/ml 5-FOA.

Figure 4.1. Pipeline of the N starvation and repletion experiment, including the list of P. tricornutum strains used, the time points analysed and the measurements performed. T0 indicates the beginning of the experiment, before the N starvation beginning, T4 indicates N starvation reached after four days in 50 µM NaNO₃, T5 indicates the day after repletion with 882 µM NaNO₃ (5 days from the beginning of the experiment) and T7 indicates three days after N repletion (7 days from the beginning of the experiment). The three flasks represent three biological replicates for each strain.

Figure 4.2. Phyre2 output showing the predicted secondary structure of the N-terminal region of PtNPF1 amino acid sequence. The di-leucine based motif representing the predicted vacuolar signal peptide is highlighted in red. The first predicted transmembrane helix (TMH 1) is indicated in gold.

Figure 4.3. Vectors generated for PtNPF1 overexpression, through co-transformation with resistance plasmid: A) Lhcf2p-PtNPF1-YFP-Lhcflτ plasmid, with PtNPF1 upstream of YFP and controlled by the P. tricornutum Lhcfl2 strong promoter; B) PmH4p-GFP-PtNPF1-Lhcflτ plasmid, with PtNPF1 downstream of GFP and controlled by the P. multistriata PmH4 constitutive promoter; C) Lhcfl6p-Sh-Ble-Lhcflτ plasmid, with Sh-Ble gene conferring Phleomycin resistance (Falciatore et al., 1999).

Figure 4.4. Selection and PCR screening of PtNPF1-YFP overexpressing strains. A) Schematic representation of the PtNPF1-YFP cassette with the two primer couples used, represented by red and green arrows and amplifying two partially overlapping regions. C) Agarose gels showing the results of the PCR amplification of fragments from promoter to the gene of interest PtNPF1 and from PtNPF1 to YFP. M represents the 1Kb marker, C+ the plasmid and C- the blank.

Figure 4.5. PtNPF1 gene relative expression levels: A) agarose gel showing good quality RNA extracted from P. tricornutum wild-type and PtNPF1 overexpressing strains 1OE 4 and 1OE 9; B) PCR performed with intron-spanning primers to check gDNA contamination (the amplicon from the gDNA containing intron is 150 bp, while the amplicon from the cDNA without intron is 90 bp). M represents the 1Kb marker. C) PtNPF1 relative gene expression levels of overexpressing strains, normalized on the internal control RPS gene and compared to wild-type set as zero. Error bars represent the standard deviation of three technical replicates of two biological replicates.

Figure 4.6. Selection and PCR screening of GFP-PtNPF1 overexpressing strains. A) Schematic representation of the GFP-PtNPF1 cassette with the two primer couples used, represented by red and green arrows and amplifying two partially overlapping regions. B) Agarose gels showing the results of the PCR amplification of fragments from promoter to GFP and from GFP to the gene of interest PtNPF1. M represents the 1Kb marker, C+ the plasmid and C- the blank.
Figure 4.7. Subcellular localisation of PtNPF1 in *P. tricornutum*. Confocal microscopy images of *P. tricornutum* cells expressing C-terminal YFP-tagged PtNPF1. First column: BF corresponds to Bright-Field; second column: chlorophyll *a* autofluorescence in red; third column: YFP fluorescence in green; fourth column: merged channels. Scale bar: 5 µm. 

Figure 4.8. Subcellular localisation of PtNPF1 in *P. tricornutum*. Confocal microscopy images of *P. tricornutum* cells expressing N-terminal GFP-tagged PtNPF1. First column: BF corresponds to Bright-Field; second column: chlorophyll *a* autofluorescence in red; third column: GFP fluorescence in green; fourth column: merged channels. Scale bar: 5 µm. 

Figure 4.9. Confocal microscopy images of *P. tricornutum* wild-type cells stained with the vacuole tracker MDY-64. First column: chlorophyll *a* autofluorescence in red; second column: MDY-64 tonoplast staining fluorescence in cyan; third column: merged channels. Scale bar: 5 µm. 

Figure 4.10. Morphological parameters and intracellular NO$_3^-$ content of *P. tricornutum* wild-type in different N concentrations. A-D) Confocal microscopy images of *P. tricornutum* wild-type cells stained with the vacuole tracker MDY-64, after 4 days in normal N conditions (A and B) and N starvation (C and D). Chlorophyll *a* in red, tonoplast tracker MDY-64 in cyan. Scale bar: 5 µm. E) number of vacuoles and F) ratio between total vacuole area and total cellular area. The values are shown as mean ± SD (N= see Supplementary Table S4.2). G) intracellular NO$_3^-$ content in *P. tricornutum* cells grown in normal N condition and N starvation. The values are shown as mean ± SD (N=3 biological replicates analysed for intracellular NO$_3^-$ content per each condition). * indicates p < 0.05. 

Figure 4.11. Confocal microscopy images of *P. tricornutum* wild-type, *Ptnpf1* knock-out mutant 2.9 and PtNPF1-YFP overexpressing strain 1OE 4 after 0, 2, 4 and 7 days of N starvation. The number of total observations is reported in the Supplementary Table S4.2. Chlorophyll *a* is shown in red, tonoplast tracker MDY-64 in cyan. For PtNPF1 overexpressing strain 1OE 4, tonoplast visualisation was performed through the fluorescence of the YFP tag, in green. Scale bar: 5 µm. 

Figure 4.12. Morphological parameters and intracellular NO$_3^-$ content of *P. tricornutum* wild-type (black), *Ptnpf1* knock-out mutant 1KO 2.9 (blue) and PtNPF1-YFP overexpressing strain 1OE 4 (green) after different days of N starvation, namely 0, 2, 4 and 7 days. A) number of vacuoles, B) ratio between total vacuole area and total cellular area and C) ratio between total chloroplast area and total cellular area. The values are shown as mean ± SD (N= see Supplementary Table S4.2). 

Figure 4.13. Intracellular NO$_3^-$ content in *P. tricornutum* wild-type, *Ptnpf1* knock-out mutant 1KO 2.9 and PtNPF1-YFP overexpressing strain 1OE 4 cells grown in 882 µM NaNO$_3$ and 882 µM NH$_4$Cl as alternative N sources. The values are shown as mean ± SD (N=3 biological replicates per each strain and condition). 

Figure 4.14. Growth curves of *P. tricornutum* wild-type, PtNPF1-YFP overexpressing strains and *Ptnpf1* mutants in different NO$_3^-$ concentrations and N sources: A-B) 882 µM NaNO$_3$, C-D) 50 µM NaNO$_3$, E-F) 882 µM NH$_4$Cl and G-H) 882 µM urea. In *vivo* chlorophyll *a* fluorescence was used for measurements as proxy for growth, and natural logarithm of growth curves was calculated. A knock-out mutant for a different gene (2KO 1.15), generated with the same proteolistic protocol, was used as additional control strain.
In all the graphs, error bars represent the standard deviation of three biological replicates. a.u. arbitrary units. * indicates $p < 0.05$; ** for $p < 0.01$ and *** for $p < 0.001$.

Figure 4.15. Growth curves of *P. tricornutum* wild-type, PtNPF1-YFP overexpressing strains and Ptnpf1 mutants during N starvation and repletion experiment: A) *In vivo* chlorophyll a fluorescence was used for measurements as proxy for growth; B) natural logarithm of the *in vivo* chlorophyll a fluorescence; C) liquid cultures used for the experiment, photographed two days after N repletion (on day 6). Labels indicate the different strains. Note the paler colour of flasks of the Ptnpf1 knock-out mutants. A knock-out mutant for a different gene (2KO 1.15), generated with the same proteolistic protocol, was used as additional control strain. In all the graphs, error bars represent the standard deviation of three biological replicates. a.u. arbitrary units. * indicates $p < 0.05$; ** for $p < 0.01$ and *** for $p < 0.001$.

Figure 4.16. Growth curves of *P. tricornutum* wild-type, PtNPF1-YFP overexpressing strains and Ptnpf1 mutants during P or different N sources starvation and repletion experiments: A) PO$_4^{3-}$ starvation and repletion, B) NO$_3^-$ starvation and repletion, and C) NH$_4^+$ starvation and repletion. *In vivo* chlorophyll a fluorescence was used for measurements as proxy for growth. A knock-out mutant for a different gene (2KO 1.15), generated with the same proteolistic protocol, was used as additional control strain. In all the graphs, error bars represent the standard deviation of three biological replicates. a.u. arbitrary units. * indicates $p < 0.05$; ** for $p < 0.01$ and *** for $p < 0.001$.

Figure 4.17. Growth and NO$_3^-$ concentrations in wild-type and Ptnpf1 knock-out cells. A) Cell counts for wild type, PtNPF1-YFP overexpressing strains and Ptnpf1 knock-out strains. A knock-out mutant for a different gene (2KO 1.15), generated with the same proteolistic protocol, was used as additional control strain. B) Measurement of *in vivo* chlorophyll a autofluorescence providing a proxy for growth. C) Extracellular NO$_3^-$ content measured by UV spectrophotometry. D) Intracellular NO$_3^-$ extracted from cells and measured by UV spectrophotometry. In all the graphs, error bars represent the standard deviation of three biological replicates. a.u. arbitrary units. * indicates $p < 0.05$; ** for $p < 0.01$ and *** for $p < 0.001$.

Figure 4.18. Relative expression levels of genes involved in N transport and metabolism in *P. tricornutum* wild-type, Ptnpf1 knock-out mutant 1KO 2.9 and Ptnpf2 knock-out mutant 2KO 1.15 strains during N starvation and repletion experiment. A-B) PtNPFs, C-D-E) PtNRT2s predicted to the tonoplast, F-G-H) PtNRT2s predicted to the plasma membrane, I-L) PtNR and PtNAR1 genes involved in N assimilation, M) PtCLC3 encoding a tonoplast chloride channel. T0 represents the beginning of the experiment, before the N starvation beginning, T4 indicates N starvation reached after four days in 50 µM NaNO$_3$, T5 indicates the day after repletion with 882 µM NaNO$_3$ (5 days from the beginning of the experiment) and T7 indicates three days after N repletion (7 days from the beginning of the experiment). Wild-type at T0 was set as reference condition at zero. Error bars represent the standard deviation of three technical replicates of two biological replicates. Differences ≥ 2 fold change, indicated by the horizontal dashed lines, are considered statistically significative.

Figure 4.19. Model of the action exerted by PtNPF1 in regulating N fluxes inside the cell after N starvation and repletion. A) PtNPF1 (represented as green protein on the tonoplast membrane) putatively involved in vacuole NO$_3^-$ sensing and signalling, together with information on cytosolic and extracellular N availability, regulates N fluxes, in concert with high-affinity NO$_3^-$ transporters (PtNRT2s, represented as pink, yellow and blue proteins on the plasma membrane and on the tonoplast), in wild-type cells. B) Ptnpf1
knock-out mutants lack PtNPF1, leading to an impaired regulation of N fluxes and, after
starvation and repletion, to a slower downregulation of PtNRT2 genes expression..... 188

Figure 5.1. Diatoms bilobate-shaped secondary plastid surrounded by four membranes.
The outermost membrane is in continuum with the host rough endoplasmic reticulum.
The space between the two inner- and outermost membrane pairs (PPC) represents the
former red algal cytoplasm of the endosymbiont. cERM: chloroplast ER membrane;
PPM: periplastidal membrane; PPC: periplastidal compartment; oEM: plastid outer
envelope membrane; iEM: plastid inner envelope membrane............................... 196

Figure 5.2. Pipeline of the shift experiment from normal pH (pH 8) to low pH (pH 7),
including the list of *P. tricornutum* strains used, the time points analysed and the
measurements performed. The three Erlenmeyer flasks represent three biological
replicates for each strain. .............................................................................................. 203

Figure 5.3. Vectors generated for PtNPF2 overexpression, through co-transformation
with resistance plasmid: A) Lhcf2p-PtNPF2-YFP-Lhcf1t plasmid, with PtNPF2 upstream
of YFP and controlled by the *P. tricornutum* *Lhcf2* strong promoter; B) PmH4p-GFP-
PtNPF2-Lhcf1t plasmid, with PtNPF2 downstream of GFP and controlled by the *P.
multistriata* *PmH4* constitutive promoter; C) Lhcf6p-Sh-Ble-Lhcf1t plasmid, with *Sh-Ble*
gene conferring Phleomycin resistance (Falciatore et al., 1999). ................................. 209

Figure 5.4. PCR screening on *P. tricornutum* colonies growing on selective medium. A)
Transformed cells growing on selective medium, with non-transformed control on
selective (C-) and non-selective (C+) plates. B) Schematic representation of the PtNPF2-
YFP cassette with the two primer couples used, represented by red and green arrows and
amplifying two partially overlapping regions. C) PCR showing amplification of a
fragment from promoter to the gene of interest PtNPF2 and a fragment from PtNPF2 to
YFP. M represents the 1Kb marker, C+ the plasmid and C- the blank......................... 210

Figure 5.5. PtNPF2 relative gene expression levels of *P. tricornutum* PtNPF2
overexpressing strains 2OE 2 and 2OE 3, normalised on the internal control RPS gene
and on the wild-type, set as zero. Error bars represent the standard deviation of three
technical replicates of two biological replicates (for a total of 6 measurements)......... 211

Figure 5.6. PCR screening on *P. tricornutum* colonies growing on selective medium. A)
Schematic representation of the GFP-PtNPF1 cassette with the two primer couples used,
represented by red and green arrows and amplifying two partially overlapping regions.
B) PCR showing amplification of a fragment from promoter to the GFP gene and a
fragment from GFP to the gene of interest PtNPF2. M represents the 1Kb marker, C+ the
plasmid and C- the blank................................................................. 212

Figure 5.7. Subcellular localisation of PtNPF2 in *P. tricornutum* through PtNPF2-YFP
fusion protein expression. First column: BF corresponds to Bright-Field; second column:
chlorophyll a autofluorescence in red; third column: YFP fluorescence in green; fourth
column: HOECST staining for nucleus visualisation in blue; fifth column: merged
channels. Scale bar: 5 µm. ........................................................................................ 213

Figure 5.8. Subcellular localisation of PtNPF2 in *P. tricornutum* through GFP-PtNPF2
fusion protein expression. First column: BF corresponds to Bright-Field; second column:
chlorophyll a autofluorescence in red; third column: GFP fluorescence in green; fourth
column: merged channels. Scale bar: 5 µm. ............................................................... 214
Figure 5.9. pPha-NR-sHsp70_BTS_mRuby3-NAT plasmid provided by Dr. Meier, containing the BTS of the gene encoding Hsp70 in frame with the mRuby fluorescent tag and the gene conferring resistance against Nourseothricin.

Figure 5.10. PtNPF2 gene schemes of *P. tricornutum* wild-type and *Ptnpf2* mutants 2KO 1.15 and 1.16, showing Cas9 different cuts on the two mutated alleles. Yellow bars represent the two gRNAs.

Figure 5.11. Growth curves of *P. tricornutum* wild-type, two PtNPF2-YFP strains and two Ptnpf2 mutants in different NO$_3^-$ concentrations and N sources: A-B) 882 µM NaNO$_3$, C-D) 50 µM NaNO$_3$, E-F) 882 µM NH$_4$Cl and G-H) 882 µM urea. *In vivo* chlorophyll a fluorescence was used for measurements as proxy for growth, and natural logarithm of growth curves was calculated. A knock-out mutant for a different gene, generated with the same proteolitic protocol (1KO 2.9), was used as additional control strain. In all the graphs, error bars represent the standard deviation of three biological replicates. a.u. arbitrary units. * indicates $p < 0.05$; ** for $p < 0.01$ and *** for $p < 0.001$.

Figure 5.12. Growth curves of *P. tricornutum* wild-type, two PtNPF2-YFP overexpressing strains and two Ptnpf2 mutants in different pH conditions: A-B) pH 9, C-D) pH 8, E-F) pH 7 and G-H) pH 6. *In vivo* chlorophyll a fluorescence was used for measurements as proxy for growth, and natural logarithm of growth curves was calculated. A knock-out mutant for a different gene, generated with the same proteolistic protocol (1KO 2.9), was used as additional control strain. In all the graphs, error bars represent the standard deviation of three biological replicates. a.u. arbitrary units. * indicates $p < 0.05$; ** for $p < 0.01$ and *** for $p < 0.001$.

Figure 5.13. Growth curves of *P. tricornutum* wild-type, two PtNPF2-YFP overexpressing strains and two Ptnpf2 mutants in medium at pH 7 supplemented with different N sources: A-B) 882 µM NaNO$_3$, C-D) 882 µM NH$_4$Cl and E-F) 882 µM urea. *In vivo* chlorophyll a fluorescence was used for measurements as proxy for growth, and natural logarithm of growth curves was calculated. A knock-out mutant for a different gene, generated with the same proteolistic protocol (1KO 2.9), was used as additional control strain. In all the graphs, error bars represent the standard deviation of three biological replicates. a.u. arbitrary units. * indicates $p < 0.05$; ** for $p < 0.01$ and *** for $p < 0.001$.

Figure 5.14. Growth of *P. tricornutum* wild-type, PtNPF2-YFP overexpressing strains and Ptnpf2 mutants. A) Photo of different wild-type and transgenic strains after four days from the shift to low pH (pH 7). Labels indicate different strains. Note the paler colour of the flasks of the Ptnpf2 knock-out mutants. B) Growth rates calculated on cell concentrations of *P. tricornutum* wild-type, two PtNPF2-YFP overexpressing strains and two Ptnpf2 mutants, calculated at days 1 and 2 on previous growth curves performed at pH 7 (Figs. 5.12E-F). Error bars represent standard deviations of three biological replicates. * indicates $p < 0.05$.

Figure 5.15. Growth curves of *P. tricornutum* wild-type, two PtNPF2-YFP overexpressing strains and two Ptnpf2 mutants after the shift to A) pH 8 and B) pH 7. Natural logarithm of growth curves was calculated, respectively for shift to C) pH 8 and D) pH 7. Growth rates were calculated at different time points for the shift to E) pH 8 and F) pH 7. Error bars represent standard deviations of three biological replicates. * indicates $p < 0.05$; ** for $p < 0.01$ and *** for $p < 0.001$. 
Figure 5.16. $F_v/F_m$ values for $P. tricornutum$ wild-type, two $PtNPF2$ overexpressing strains and two $Ptnpf2$ mutants after the shift to normal pH and low pH. A) $F_v/F_m$ values after 3, 24 and 48 hours from the shift to pH 8. B) $F_v/F_m$ values after 3, 24 and 48 hours from the shift to pH 7. Error bars represent standard deviations of three biological replicates. * indicates $p < 0.05$; ** for $p < 0.01$ and *** for $p < 0.001$. 

Figure 5.17. $rETR_{max}$ values for $P. tricornutum$ wild-type, two $PtNPF2$ overexpressing strains and two $Ptnpf2$ mutants after the shift to normal pH and low pH. A) $rETR_{max}$ values after 3, 24 and 48 hours from the shift to pH 8. B) $rETR_{max}$ values after 3, 24 and 48 hours from the shift to pH 7. Error bars represent standard deviations of three biological replicates. * indicates $p < 0.05$ and *** for $p < 0.001$. 

Figure 5.18. NPQ$_{max}$ values for $P. tricornutum$ wild-type, two $PtNPF2$ overexpressing strains and two $Ptnpf2$ mutants after the shift to normal pH and low pH. A) NPQ$_{max}$ values after 3, 24 and 48 hours from the shift to pH 8. B) NPQ$_{max}$ values after 3, 24 and 48 hours from the shift to pH 7. Error bars represent standard deviations of three biological replicates. * indicates $p < 0.05$ and *** for $p < 0.001$. 

Figure 5.19. NPQ half-time for relaxation ($T_{50}$) values for $P. tricornutum$ wild-type, two $PtNPF2$ overexpressing strains and two $Ptnpf2$ mutants after the shift to normal pH and low pH. A) $T_{50}$ values after 3, 24 and 48 hours from the shift to pH 8. B) $T_{50}$ values after 3, 24 and 48 hours from the shift to pH 7. Error bars represent standard deviations of three biological replicates. * indicates $p < 0.05$; ** for $p < 0.01$ and *** for $p < 0.001$. 

Figure 5.20. Relative intracellular pH ($pH_i$) values for $P. tricornutum$ wild-type, two $PtNPF2$ overexpressing strains and two $Ptnpf2$ mutants after the shift to normal pH and low pH. $pH_e$ represents the pH of the external medium. Error bars represent standard deviations of three biological replicates. * indicates $p < 0.05$; ** for $p < 0.01$ and *** for $p < 0.001$. 

Figure 5.21. Photosynthetic parameters for $P. tricornutum$ wild-type, two $PtNPF2$ overexpressing strains and two $Ptnpf2$ mutants before and after the shift from dark to light. A) Cell concentration, B) maximum fluorescence $F_{max}$, C) $F_v/F_m$, D) $\alpha$, E) $rETR_{max}$ and F) NPQ$_{max}$ values before the shift (dark) and 3 and 24 hours after the shift to light. Error bars represent standard deviations of three biological replicates. * indicates $p < 0.05$. 

Figure 5.22. Rates of net photosynthesis as a function of HCO$_3^-$ concentration for $P. tricornutum$ wild-type, two $PtNPF2$ overexpressing strains and two $Ptnpf2$ mutants. Error bars are not present as these are preliminary results representing only one replicate. 

Figure 5.23. RNA samples preparation for RNA-seq analysis. A) $P. tricornutum$ wild-type and $Ptnpf2$ knock-out mutant 2KO 1.15 growing in flasks after 24 hours from the shift to low pH (pH 7). Labels indicate different strains. B) RNA extracted from cells and run on 1% agarose gel. Two samples (WT pH 8 samples 1 and 3) were re-extracted because of the low RNA amount of the initial samples. M represents the 1Kb marker. C-D) PCR on $PtNOA$ gene to check genomic DNA contamination in RNA samples before (C) and after DNase I treatment (D). 150 bp band indicates genomic DNA, 90 bp band indicates intron-less cDNA, here used as positive control of the PCR, and no band indicates absence of DNA in the preparation. M represents the 100 bp marker, marker, as control for the gel quality.
Figure 5.24. Electron micrograph scanning of a *Phaeodactylum* PPC. A) Original micrograph and B) organelles and compartment reconstruction. The outer nuclear envelope is shown in blue and it is in continuity with the cERM. The VN fills the space between the PPM and oEM, shown in light green. C, chloroplast; M, mitochondrion; N, nucleus; Thyl, thylakoids. Image adapted from Flori et al. (2016)............................... 239

**List of Tables**

Table 1.1. Genetic features of established and emerging diatom model species. Table adapted from Falciatore et al. (2020).......................................................................................... 19

Table 1.2. Technologies and genome-enabled tools available for diatoms. Adapted from Falciatore et al. (2020)......................................................................................................... 27

Table 2.1. List of primers used in qPCRs. ................................................................. 66

Table 2.2. List and properties of diNPFs, modified and integrated from Rogato et al. (2015). Superscript letters in the protein ID column denote allelic pairs. The *F. cylindrus* 256377 sequence was incomplete and was not analysed more in detail. TMs: transmembrane domains.................................................................................................. 70

Table 2.3. *diNPFs* gene expression data from literature.............................................. 82

Table 2.4. NO$_3^-$ regulator motifs in upstream *diNPF* sequences. ............................ 88

Table 2.5. Properties of PtNPFs and templates used to build their structural homology models. Pdb: Protein Data Bank ID. ................................................................................. 90

Table 2.6. Conserved residues involved in proton or substrate binding and transport. 92

Table 2.7. Predicted subcellular localisation of diNPFs, including sequence number and corresponding percentage that are predicted to a specific subcellular localisation, and homologous Eukaryota and bacteria proteins that drove these predictions. * indicates that LocTree3 does not find a homologous sequence whose exact location is known, but based on the homology with different sequences, the algorithm is however able to reconstruct the path that leads a protein to locate in a given compartment. ...................................... 97

Table 3.1. List of crRNAs designed for CRISPR/Cas9 proteolistic transformation, with the PAM sequence underlined. ..................................................................................... 111

Table 3.2. Mix of components for each RNP complexes to assemble, specific for the transformation experiments performed. For each experiment, the total Cas9 amount has been calculated.......................................................... 113

Table 3.3. List of primers used to amplify target genes fragments, in order to test designed gRNAs......................................................................................................................... 115

Table 3.4. List of primers used for the screening, amplifying fragments on the selective gene *PtAPT* and on the target genes *PtNPF1* and *PtNPF2*.................................................................. 117

Table 4.1. List of primers used for plasmid construction.................................................. 156

Table 4.2. List of primers used for overexpressing mutants screening............................ 158
Table 4.3. List of primers used for qPCRs................................................................. 162
Table 5.1. List of primers used for plasmid construction......................................... 199
Table 5.2. List of primers used for overexpressing mutants screening..................... 200
Table 5.3. List of primers used for qPCRs................................................................. 201
Chapter 1: Introduction
1.1. Diatoms

1.1.1. Life in the Ocean

Our planet as viewed from space is mainly blue. This is because it is largely covered by water, in particular by oceans. So why call this planet Earth? Not only the area covered by oceans is larger than that covered by land, but while terrestrial life extends from just beneath the soil surface, life in the oceans extends to their deepest 11 000 meters (Beer, 2014), and marine primary productivity may be even higher than the terrestrial one: Woodward (2007) estimated the global marine primary production to be 65 Gt carbon (C) year\(^{-1}\) while the terrestrial one was 60 Gt year\(^{-1}\).

There are different hypotheses about where life on earth originated: organic compounds led to living cells in shallow rainwater ponds, or in a mist, called “organic haze”, generated as gases such as (N\(_2\)) mixed with the condensing water vapour, that then precipitated into the primordial ocean when the planet cooled (Beer, 2014), or maybe in deep-sea hydrothermal vents (Martin et al., 2008). In any case, life originated thanks to water about 1 billion years after its formation, i.e. \(\sim 3.4\) billion years ago: fossil evidence suggests that the first life forms were “bacteria-like” life forms, similar to those found in Western Australia (Wacey et al., 2006). Still now, the majority of life underwater is represented by microbes, with around 70% of the marine biomass being characterised by bacteria and protists (Bar-On et al., 2018), that can be heterotrophic or photosynthetic.

1.1.2. Phytoplankton

Photosynthesis is one of the most important processes in nature as it sustains aerobic life on Earth and keeps the planet alive. In fact, after the origin of life, the rise of unicellular photosynthetic organisms represented a strongly revolutionary event, completely changing the Earth face.

Ancestors of cyanobacteria started generating oxygen (O\(_2\)) by oxidizing water about 2.4 billion years ago (Benoiston et al., 2017; Soo et al., 2017) (Fig. 1.1). They evolved the capability of converting carbon dioxide (CO\(_2\)) into organic carbon, chemical energy and O\(_2\) using sunlight as source of energy. In this way, the rise of unicellular photosynthetic organisms caused an O\(_2\) accumulation in the atmosphere, known as the “great oxygenation event”, which changed its redox state and promoted the proliferation of aerobic respiration (Soo et al., 2017) (Fig. 1.1). The success of oxygenic photosynthesis
greatly increased the primary productivity of the earth, sustaining aerobic life on Earth and keeping the planet alive (Fischer et al., 2016).

Following the evolution of oxygenic cyanobacteria, it took roughly 2 billion years for sophisticated multicellular animal life to arise. During this time, eukaryotic organisms appeared bearing the first mitochondria derived from the endosymbiosis of a proteobacterium in an Archaean-like cell, in which respiratory processes could occur (Benoiston et al., 2017; Martin and Russell, 2003). Subsequently, around 1.2 billion years ago, the evolution of chloroplasts was coupled with the invasion or engulfment of a cyanobacterium into the prototypical eukaryote (Fig. 1.1). However, today the forms that predominate in the ocean are primarily descended from additional or "secondary" endosymbiotic events in which eukaryotic green or red algae were incorporated into a eukaryotic cell for a second time (Benoiston et al., 2017; Falkowski et al., 2004a).

This heterogeneous group of photosynthetic microorganisms, both prokaryotic and eukaryotic, is called phytoplankton, and it is mainly composed by cyanobacteria, diatoms, dinoflagellates and coccolithophores. Although phylogenetic diversity within the phytoplankton is deeply branching, many fundamental metabolic pathways are remarkably conserved: this combination of deep genetic diversity and high functional redundancy has helped to ensure the continuity of oxygenic photosynthesis despite considerable changes in Earth climate (Falkowski et al., 1998).
1.1.3. Diatom place in the world

Diatoms are unicellular photosynthetic eukaryotes, with a significant impact in global biogeochemical cycles and in food webs as they contribute to around 20% of global primary productivity (Falkowski et al., 1998; Field et al., 1998). Diatoms are considered one of the most diverse and ecologically important phytoplanktonic groups, responsible for the fixation of about 26 Gt of carbon, which corresponds to almost half of the total fixation of organic matter in the oceans (Malviya et al., 2016; Tréguer et al., 2018) and is comparable to the combined productivity of all terrestrial forests (Field et al., 1998).

They belong to the class Bacillariophyceae, and more than 200 genera and about 100 000 species can be found on the earth (Mann and Vanormelingen, 2013). They are widely distributed in almost all aquatic habitats and can also occur as endosymbionts in dinoflagellates and foraminifers (Round et al., 1990).

Diatoms are the most abundant class composing marine biomass. This was confirmed from decades through morphological studies (Leblanc et al., 2012) and in the last years with the advent of omics technologies: metabarcoding survey based on the V9 hypervariable region of 18S rDNA, performed as part of the TARA Oceans global plankton sampling campaign, highlighted that diatoms are the most abundant and diverse group of obligate photosynthetic eukaryotes and the fifth most abundant group of marine eukaryotes (Benoiston et al., 2017; de Vargas et al., 2015).

They typically dominate well-mixed coastal and upwelling regions, where the organic C they generate supports productive fisheries, and they are often key components of spring blooms, as they can divide more rapidly than other organisms when they find favourable environmental conditions (Benoiston et al., 2017). But diatoms are also well adapted to survive for long periods in conditions of nutrient and light limitation, for example in polar oceanic environments such as Arctic and Antarctica where extreme conditions do not allow a lot of other organisms to survive (Armbrust, 2009).

Larger species of diatoms can move up and down through the water column by controlling their buoyancy, while other open-ocean species can move between nutrient-depleted but well-lighted surface waters, where they can photosynthesise, and nitrate-rich waters about 100 meters deep, where they take up and store the required nutrients (Villareal et al., 1999; Armbrust, 2009).
1.1.4. Diatom morphology and structure

Their name comes from the Greek *diá-tom-os* which literally means “to cut in half” and refers to the shape of their silica cell wall, which externally divides the cell into two valves (Armbrust, 2009). More precisely, this word is referred to the morphological features of the frustule, which is a cell wall made up of amorphous silica and consists of two overlapping valves termed thecae, which encapsulate the cell (Armbrust, 2009). Thecae are slightly unequal parts, like a box (hypotheca) with its lid (epitheca) (Fig. 1.2). As the diatom divides, each daughter retains one theca of the original frustule and produces one new theca. Both epitheca and hypotheca are constituted of several parts: the valves, similar to plates, called epivalve and hypovalve respectively, and one or several girdle bands (Benton and Harper, 2009; Fuhrmann et al., 2004). The cell wall of valves and girdles is characterised by different patterns of lines (striae), consisting of very regular arrays of chambers (areolae) containing small pores (cribra), which are believed to proceed via self-organised phase separation (Fuhrmann et al., 2004) (Fig. 1.2).

Based on the frustule symmetry, diatoms can be divided into centrics and pennates: centric diatoms, mainly planktonic, have circular valves with radial symmetry, while pennate diatoms, mainly benthic, have elongated valves with bipolar symmetry of the frustule (Kooistra et al., 2009).

The frustule has different important roles, besides being an important classification feature, among which the main ones are acting as a protective shield against grazing and acting as a photonic crystal which focuses light while filtering detrimental UV radiation. In fact, it has been demonstrated that nanostructure and optical properties of diatom frustules can influence incoming light, being useful in applications such as solar panels, which increases the energy absorption of thin-film solar cells (Chen et al., 2015).

Every mitotic division, the cell stretches by pushing the two thecae away from each other, after that each daughter cell begins to generate the other theca through a biomineralization process. The most well characterised genes involved in biomineralization encode the Silicic acid Transporters (SITs), which transport silicic acid from seawater into the cell (Durkin et al., 2016; De Tommasi et al., 2017). After uptake, silica precipitation and polymerization occur through the polarized production of Silica Deposition Vesicles (SDVs), that contain silica forming organic components. Here silica formation takes place and includes different proteins and molecules working as cell wall scaffold, such as silacidins which play a structural role in diatom frustules, Long-Chain Polyamines.
(LCPAs) which are linear polyamines constituting the main organic and soluble fraction of the biosilica matrix and, in some species, chitin which contributes to cell rigidity (De Tommasi et al., 2017).

Besides their frustule morphology, diatom cell biology is a critical and still little explored aspect. Previous work suggested that, for instance, diatoms physiology and metabolism are determined by their peculiar cell organisation and more specifically by the morphology and arrangement of key energy-producing organelle machineries, such as the plastids and the mitochondria (Bailleul et al., 2015; Maier et al., 2022).

The nucleus of diatoms is surrounded by a typical nuclear envelope with pores, like other eukaryotes. The outer membrane of the nuclear envelope is connected to the endoplasmic reticulum membrane but it is also continuous with the outermost membrane of the plastid (Flori et al., 2016; Cavalier-Smith, 2000).

Like most plant and algal cells, diatoms also harbour a vacuole, which can store different types of molecules (Fig. 1.2). It can store nutrients to be reallocated during unfavourable conditions, possibly being one of the mechanisms defining diatoms ecological success over fluctuating nutrient availability in the ocean (Behrenfeld et al., 2021), but it is also involved in osmotic pressure maintenance and buoyancy regulation (Behrenfeld et al., 2021).

Figure 1.2. Schematic representation of diatom morphology. On the left, representation of frustule morphological features. On the right, organelles organisation inside diatom cell.
The mitochondria of diatoms vary in shape and quantities and, like other eukaryotes, they possess two membranes, of which the innermost one forms tubular cristae (Maier et al., 2022) (Fig. 1.2). On the other hand, plastid types found in diatoms are restricted to most organisms of the supergroup Chromalveolata, they contain four surrounding membranes, originated through a process of secondary endosymbiosis (see below paragraph 1.1.6 “Diatoms as result of secondary endosymbiosis”). As in other microalgae, the diatom plastid is organised in such a way that the electron transport reactions and the CO$_2$-fixing reactions of photosynthesis occur in two distinct compartments: the thylakoids, which are not organised in grana as plants but form stacks of three layers called girdle lamellae (Fig. 1.2), host the proteins performing the light-phase (electron transport reactions), and the pyrenoid, which occupies the centre of the plastid and is crossed by special thylakoids, forming a Rubisco-rich compartment for CO$_2$ fixation (Maier et al., 2022; Flori et al., 2016).

1.1.5. Diatom evolution

It is deemed that diatoms arose on earth in the Triassic period, 250 million years ago even if the first good-preserved diatom fossil dates back to early Jurassic (Fig. 1.3). They probably became dominant on the planet about 135 million years ago (Armbrust, 2009; Falciatore et al., 2020). They were able to survive to mass extinctions and, in contrast to dinoflagellates and coccolithophores, diatom diversity continued to incredibly increase from 34 to 5 million years ago (Benoiston et al., 2017). The rise of diatoms in that period was accompanied by the establishment of the main petroleum source rocks, derived from carbon export. The spatial correspondence of silica and fossil fuels, together with biomarker surveys in sediments and source rocks, indicate a crucial role of diatoms in the formation of today reserves (Benoiston et al., 2017; Cermeño, 2016).

Moreover, molecular phylogenetic dating studies (Sorhannus, 2007) helped to infer the time frame within which the diatoms originated and diversified. As described before, diatoms are broadly divided into two main groups, centrics and pennates, depending on the symmetry of their frustule. Bipolar and multipolar centric diatoms originated about 150 million years ago while pennate diatoms evolved about 96 million years ago (Sorhannus, 2007) (Fig. 1.3). Centric diatoms are further divided into polar centrics (Mediophyceae) and radial centrics (Coscinodiscophyceae), while pennates are further divided into araphid (Fragilariophyceae) and raphid (Bacillariophyceae). The oldest araphid pennates arose around 93 million years ago, while the more recent raphid pennate
Diatoms evolved about 75 million years ago (Sorhannus, 2007; Armbrust, 2009) (Fig. 1.3).

Raphid pennate diatoms are shown to have the potential for active gliding motility, enabled by a longitudinal slit in their valves called raphe, from which polysaccharides are secreted and that allows cells to slide on the substrate or to adhere on it. The raphe allowed diatoms to colonise a wide range of new benthic habitats (Sims et al., 2006). Diatom adhesion has important consequences, since diatoms are prominent biofouling organisms that can proliferate even on advanced hydrophobic anti-fouling surfaces (Callow and Callow, 2011). Moreover, due to its extraordinary adhesive properties and economic implications, the carbohydrate and glycoprotein composition of the raphid adhesive exopolysaccharides has been receiving increased attention for biotechnological applications.

Figure 1.3. Estimated timing of divergence of the four major diatom lineages and coincident events in Earth history. Maps are palaeographic reconstructions of continent locations during the emergence of the diatom lineages. Shallower depths in the ocean are indicated by lighter blues. Adapted from Armbrust (2009).
1.1.6. Diatoms as result of secondary endosymbiosis

If sediments can tell diatom temporal history, diatom genomes can give information about their evolutionary history. Already prior to the advent of genome sequencing, morphological and ultrastructural data suggested that diatoms were derived from a secondary endosymbiotic event involving an organism belonging to the Rhodophyta lineage, previously evolved in a primary endosymbiosis event (Cavalier-Smith, 2000). This event, occurred between 1200 and 700 million years ago, was common to the whole eukaryotic supergroup of Chromalveolata that includes dinoflagellates, coccolithophores and stramenopiles, the phylum to which diatoms belong (Benoiston et al., 2017; Parker et al., 2008).

Although the analysis of diatom genome (see below paragraph 1.2.1 “Omics resources”) provided support for a red algal endosymbiont, recent studies, uncovering the abundance of genes apparently derived from green algal sources, proposed the controversial hypothesis that a green algal endosymbiont preceded the red alga (Benoiston et al., 2017) (Fig. 1.4): in this scenario many genes coming from the green alga were retained before the arrival of the second one, while the red algal genes that were acquired later were not (Deschamps and Moreira, 2012; Moustafa et al., 2009) (Fig. 1.4). In this way, green and red algae genes present in diatom genomes may have led to a selective advantage in oceanic environments, and this would explain why these organisms are able to dominate in the ocean with high variable and fluctuating environmental conditions, while photosynthetic organisms hosting plastids derived from green algae only dominate terrestrial habitats (Falkowski et al., 2004a).
Figure 1.4. Schematic representation of diatom evolution. The “melting pot” diatom genome finds its origins in successive gene transfers following endosymbiosis between red and green algae and a host heterotrophic cell (Moustafa et al., 2009; Benoiston et al., 2017). The event is proposed to have permitted gene transfer from prey nuclei to host nucleus as well as from organelles to nucleus. Diatoms also seem to have acquired genes through lateral gene transfer both before and after the diversification of pennates and centrics. Image adapted from Bowler et al. (2010).

Moreover, different other mechanisms contributed to the generation of diatom diversity, during the reduction of the endosymbiont: i) large-scale within-genome duplication events; ii) differential losses and gains of genes and introns, mainly in centric diatoms; iii) Horizontal Gene Transfer (HGT), represented by the presence of several hundreds of
bacterial genes scattered throughout diatom genomes; and iv) transposable elements (TEs) mobilisation (Armbrust et al., 2004; Bowler et al., 2008; Mock et al., 2017; Osuna-Cruz et al., 2020; Tanaka et al., 2015; Traller et al., 2016).

In particular, diatom genomes have small gene sizes and intergenic regions, with a significant proportion of TEs which are predominantly of the long-terminal-repeats (LTR) retrotransposon superfamily (Maumus et al., 2009; Tirichine et al., 2017). Examination of LTR retro-elements in both *P. tricornutum* and *T. pseudonana* revealed the existence of seven groups of diatom-specific TEs named CoDi (Copia-like elements from diatoms), expressed under specific conditions and suggesting a role of TEs in adaptation and diversification of diatoms (Maumus et al., 2009; Tirichine et al., 2017).

In addition, different transfer events appear to have occurred with prokaryotes at different points in diatom evolutionary history (Fig. 1.4). HGT, also called lateral gene transfer, is the transfer of genetic information between different species by a process of direct passage and integration of genetic material. In diatoms, 36 bacterial HGT events contributed to an average of 3-5% of the genetic pool through multiple transfer events spanned over time and from different prokaryotic donors (Vancaester et al., 2020). For example, 10 of the 15 gene transfers involving members of the *Deinococcus-Thermus* clade encode metabolic processes including diatom growth in the absence of vitamin B12, symbiosis and host pathogen interactions (Rastogi et al., 2018). Moreover, the integration of viral genomic material, although observed, occurred to a lesser extent in diatom genomes than what was observed in other eukaryotes (Hongo et al., 2021).

All these phenomena, revealed by several diatom species genome sequencings, contributed to generate the rich diversity of diatom species, with a high degree of metabolic flexibility that has played a major role in determining their success in contemporary oceans (Rastogi et al., 2018).

The chimeric diatom genomes contain unique combinations of genes that collectively encode non-canonical pathways of nutrient assimilation and metabolite management. An example of their chimeric nature is represented by nitrogen (N) metabolism, which is a combination of plant-like, animal-like, and bacteria-like traits. The nitrate (NO$_3^-$) assimilation pathway is similar to that of plants, indicating that diatoms acquired the ability to assimilate NO$_3^-$ from the endosymbiont (Robertson, 2006), but it is surprising to find that diatoms possess a metazoan-like urea cycle (Armbrust et al., 2004). In animals, the urea cycle is necessary to excrete N in excess, while in diatoms it has been
proposed to have a role in redistribution of central N-containing metabolites during fluctuating N availability (Allen et al., 2011; Smith et al., 2019).

1.1.7. Diatom cell cycle

Living in a rigid glass box has several limitations, which influence diatom life cycle. Diatoms have a diplontic life cycle, which includes the diploid phase that is the vegetative cell and the haploid phase that is represented by gametes. As mentioned before (see paragraph 1.1.4 “Diatom morphology”), during mitotic division, the two daughter cells keep one maternal theca that becomes the larger epitheca and synthesise ex novo the new smaller hypotheca, regardless of whether it inherited the parental hypotheca or epitheca (Montresor et al., 2016). The new hypotheca of the daughter cell is produced within the confines of the maternal frustule in few hours and derive from the fusion of SDVs with cell plasma membrane; the deposition starts immediately after nuclear mitotic division. This implies that the two daughter cells differ in size and, as cell divisions proceed, the cell size range progressively widens, while the mode of the cell-size distribution of the population decreases (Fig. 1.5): this phenomenon called MacDonald-Pfitzer rule predict that, within a population, mitotic divisions will decrease the mean apical length and will increase the variance in length (Macdonald, 1869; Pfitzer, 1869).

To avoid cell death through miniaturisation, cells reproduce also sexually. In fact, most diatoms have a life cycle in which there is an alternation between long periods (months, years) of vegetative growth involving mitotic cell division, and a short period (hours, days, rarely 1 month) in which sexual reproduction takes place (Seckbach, 2019). Centrics and pennates produce gametes with different morphology (Fig. 1.5): the gametes of centrics have unequal size and morphology with large egg cells and small and motile sperm cells (anisogametes), while the gametes of pennates are morphologically similar (isogametes) (Fig. 1.5).

Genetic determination of diatom sex seems to be driven by a single sex locus for which one mating type is generally heterozygous (Bileke et al., 2022). The first identification of a genetic sex-determining locus was made in Seminavis robusta (Vanstechelman et al., 2013), while five mating type-related genes that show mating type-specific expression (three MT+ and two MT- specific genes) were later identified in Pseudo-nitzschia multistriata by Russo et al. (2018a).
A distinctive feature of the diatom sexual phase is that sex cannot occur over the whole cell size range: only cells below a species-specific size threshold can become sexualised (Fig. 1.5). Spontaneous and experimentally induced abrupt cell size reduction demonstrates that sexualisation is size-dependent rather than based on population age or the number of cell divisions (Bilcke et al., 2022).

Thanks to sexualisation, a new zygote not surrounded by a rigid siliceous frustule develops and it can thus expand forming the auxospore, within which the cell restores the maximum species-specific size (Fig. 1.5).

Figure 1.5. Schematic drawing of the life cycle of a centric and a pennate diatom. Diatom cells are diploid and are surrounded by a rigid frustule made of two unequal thecae. During mitosis, the new thecae are formed inside the frustule, causing a progressive decrease in the population cell size. Once a species-specific size threshold is reached, the formation of gametes takes place. Conjugation of the haploid gametes produces a zygote that expands into an auxospore, from which the large initial cell is synthesised.

Some species do not show evidence of cell size reduction and a sexual phase has not yet been observed, such as in *Thalassiosira pseudonana* (Hildebrand et al., 2006) and *Phaeodactylum tricornutum* (De Martino et al., 2007). Both species lack a cell size reduction-restoration cycle, and there are no conclusive observations of sexual reproduction despite recent indirect indications from both species that sex might be possible (Bilcke et al., 2022; Koester et al., 2018).
Moreover, several planktonic species are known to form resting stages, including spores or resting cells. While spores are morphologically distinct from vegetative cells, resting stages have valves, that are very different from their vegetative counterparts and are characterised by a condensed protoplast (Kaczmarska et al., 2013; Bilcke et al., 2022). Resting stages have a strongly reduced metabolism, they may sink to the bottom of the sea and accumulate in the sediments, constituting a reservoir analogous to the seed banks of higher plants, and are implicated in the initiation of planktonic diatom blooms. They were shown to successfully germinate even after 100 years or more of dormancy (Bilcke et al., 2022) when environmental conditions are not optimal for growth, so remaining viable in sediments for a long time (Pelusi et al., 2020).

1.1.8. Ecological contributions of diatoms

Large-scale patterns of diatom distributions from field observations and more recently from remote sensing reveal their importance as major component of phytoplankton biomass at high latitudes during spring (April-June in the Northern Hemisphere and October-December in the Southern Ocean) and in equatorial and coastal upwelling regions (Tréguer et al., 2018; Benoiston et al., 2017) (Fig. 1.6). Moreover, symbiotic diatom-diazotroph assemblages (DDAs) appear to allow diatoms to periodically be important components of plankton communities in oligotrophic regions, while other diatom species appear well suited to thrive at the sea-ice edge (Foster and Zehr, 2006; Tréguer et al., 2018). These large-scale diatom distribution patterns are captured by ecosystem models, relating phytoplankton traits with different physico-chemical parameters (Dutkiewicz et al., 2015a). A small-scale heterogeneity in diatom distribution has been shown too, particularly in relation with mesoscale eddies and sub-mesoscale fronts (Tréguer et al., 2018). This distribution of diatoms is tightly coupled with the physical variables of the ocean such as the supply of nutrients, but also with the biological ones such as the interaction with predators, pathogens or symbionts.
Figure 1.6. Extent of diatom-rich sediments compared with the distribution of modern diatoms in the ocean. Biosiliceous sediments are present in regions that, still today, are largely dominated by diatoms, in particular the Southern Ocean. A) Small dots represent seafloor sediment samples defined as containing predominantly diatom sediments, muds, also mixed with calcareous one. These data come from the Index to Marine and Lacustrine Geological Samples (Curators of Marine and Lacustrine Geological Samples Consortium, 2014; Dutkiewicz et al., 2015b). Circles of varying size and blue color correspond to diatom relative abundances determined by the Marine Ecosystem Model Inter-Comparison Project (Leblanc et al., 2012) and by the TARA Oceans survey (Malviya et al., 2016), based on metabarcoding data coming from the photic zone (both subsurface SRF and deep chlorophyll maximum DCM) and corresponding to four size classes (0.8–5 μm, 5–20 μm, 20–180 μm, and 180–2 000 μm). B) Water column inventory of diatom biomass (mmol C/m²) from a biogeochemical/ecosystem simulation (Dutkiewicz et al., 2015a). Image adapted from Benoiston et al. (2017).

As a consequence of their ecological success, their fast growth rates and relatively large sizes, diatoms are very important contributors in biogeochemical cycles. They primarily contribute to the C cycle, and in particular to the C export from the surface ocean to depth as part of the biological pump (Tréguer et al., 2018). Diatoms vary widely in size, morphology and elemental composition, which control the quality, quantity and sinking speed of biogenic matter to depth. Moreover, their silica shells provide ballast to marine snow and fecal pellets, helping C transport to both the mesopelagic layer and deep ocean.
Also life cycles and strategies, such as the chain formation or resting stages, modulate their contribution to the biological pump (Tréguer et al., 2018).

Furthermore, diatoms participate to the biogeochemical cycles of nutrients, such as nitrogen (N) and silicon (Si). Their contribution usually starts from their uptake in the cell and follows in their introduction in the food web through upper trophic levels or alternatively their sedimentation via a sinking down process (Armbrust, 2009). So, changes in both phytoplankton communities can have effects throughout the food web. This, in turn, influences the lability of particles, their sinking and remineralisation, and the transfer of carbon to higher trophic levels (Tréguer et al., 2018).

Apart from biogeochemical cycles, diatoms ecological significance is partly due to their capacity to form high biomass “blooms” following exponential cell division in response to favourable environmental conditions (Armbrust 2009). These blooms are a key source of organic carbon to marine food webs and can lead to significant carbon export from surface water (Rynearson et al., 2022). Nutrient availability, together with light and temperature, are primary determinants of diatom biomass accumulation: in fact, a reported increased in bloom events has been shown to be linked to nutrient availability, often related to anthropogenic eutrophication or coastal upwelling events (Kudela et al., 2010).

This idea fits in nicely with the well-known Margalef's Mandala, recently revisited and extended (Glibert, 2016), which depicts the seasonal shift from diatoms in spring to dinoflagellates in summer along a gradient of decreasing turbulence and nutrient availability (Margalef, 1979, 1978). It suggests an upwelling-dominated species succession which fits both with the apparent scarce ability of dinoflagellates to sequester nutrients at low concentrations, and with the strong ability of diatoms to cope with large nutrient fluctuations (Kudela et al., 2010; Glibert, 2016). In fact, compared to other phytoplanktonic groups, diatoms are generally characterised by higher maximum uptake rates of nutrients and by their ability in nutrient assimilation and storage, allowing them to be the first to take advantage of sudden nutrient availability (Dutkiewicz et al., 2015a).

Diatoms blooms can have ecological implications. In particular, diatoms can be benign or harmful, forming Harmful Algal Blooms (HABs), which are fascinating but dangerous events. Around 30 species of diatoms have been identified as harmful to either fisheries, wildlife or people in health or economic terms (Fryxell et al., 2003; Villac et al., 2010). They can damage physically harming other organisms due to their cell morphology or
high biomass accumulation or producing toxins, such as domoic acid produced by *Pseudo-nitzschia*, a toxin involved in amnesic shellfish poisoning (Bates et al., 1989). Some species are also able to produce toxic secondary metabolites, often in response to the presence of grazers (Tammilehto et al., 2015).

### 1.1.9. Economic and biotechnological importance of diatoms

Aside from playing a central role in the global ecosystem, diatoms influence human society and its daily activities. The fossilisation of siliceous empty frustules of diatoms results in massive sedimentary deposits called diatomite (Ghobara et al., 2019). Diatomite is a porous rock, which was used as a lightweight building material in ancient Egypt and Turkey. It was famously mixed with nitro-glycerine creating the stable explosive dynamite (Wisniak, 2011). Diatomaceous deposits are also important for being one of the main component of fossil fuels: burial of dead diatom biomass over millions of years contributed to generate present petroleum reserves (Falkowski et al., 2004a). In particular, widespread accumulation of diatomaceous source rocks during the Oligocene and Miocene is remarkable in Pacific region, Caspian and North Sea, accounting for over 10% of world conventional oil and gas discovered so far (Cermeño, 2016).

Today, diatomite is widely used as a filtering agent (Ghobara et al., 2019). Moreover, diatomite is used as thermal insulator, soil amendment, food additive, chemical-free insecticide and as abrasive (Ghobara et al., 2019).

Diatom frustules have specific porosity, optical, and mechanical properties, which in addition to their micro and nano-fluidics allowed diatom nanotechnology development (Fig. 1.7). Several applications of diatoms optical properties have been developed, such as bacterial sensors (by immobilisation of the biosilica encapsulated bacterial cells on a sensor chip), core-shell nanomaterials for nanomedicine applications, and 3D cell printing as promising materials for regenerative medicine (Mcheik et al., 2018). Frustules from diatoms have been investigated for their potential in biosensing, by exploiting their photoluminescence properties for gas detection, and in energy field (Fig. 1.7). For example, in solar cells technologies, the ability of light trapping by diatoms valves could enhances light harvest increasing solar cells efficiency (Chen et al., 2015).

Diatoms have also great application in environmental technologies, for example for wastewater treatment (Fig. 1.7): the excess of industrial waste discharged in the aquatic system can be used as nutrient supply by diatoms (Sharma et al., 2021b). Moreover,
Diatoms own metal-binding peptides known as phytochelatins (PCs) that protect photosynthetic organisms from heavy metals. For this reason they are used in degradation of waste coming from industries working with chemicals and releasing high amount of heavy metals (Sharma et al., 2021b).

The potential of diatoms to accumulate high lipids and varied compositions of fatty acids is also important for biofuel industry (Fig. 1.7): biodiesel productivity, land use, and oil yield support the use of microalgae for commercial production compared to corn and other food crops (Sharma et al., 2021b). In fact, the oil yield for microalgae with high oil content is almost 15-fold more as compared to corn, while the land use for corn and maize is 66-fold more as compared to microalgae (Sharma et al., 2021b). However, there are still different limitation for the large-scale implementation of these applications, mainly due to their high costs of production (Santin et al., 2021a).

Finally, diatoms have an immense nutritional value that can be used to produce compounds of nutraceutical interest, such as antioxidants, animal feed and food supplements (Fig. 1.7). In fact, diatoms synthesise several photosynthetic pigments, in particular carotenoids which are powerful antioxidants, and significant amount of Polyunsaturated Fatty Acids (PUFAs) such as omega-3 that can be used as nutritional food in human diet and animal feed (Sharma et al., 2021b; Santin et al., 2021a).

Figure 1.7. Scheme of the different uses of diatoms for green industry, from the application of frustule optical properties to the environmental technologies and the different uses of metabolites produced by diatoms. Image inspired to Sharma et al. (2021b).
1.2. Molecular resources

1.2.1. Omics resources

Interest in the ecological and economic roles of diatoms led to the need to develop techniques that can be used to study and manipulate diatom genes. So, the ever faster development of omics resources like genome- and transcriptome-based techniques and molecular tools like genetic transformation are making possible to answer complex questions about diatom biology using functional genomics.

*T. pseudonana* and *P. tricornutum* are the most advanced diatom model species for molecular studies, belonging to the centric and pennate diatom groups, respectively. Their genomes were the first to be sequenced (Armbrust et al., 2004; Bowler et al., 2008) (Table 1.1), then many transcriptomic and proteomic data were produced facilitating genome annotations and gene discovery (Falciatore et al., 2020). Both species (less than 30 µm in size) have small genomes, on the order of 30 Mb. Nonetheless, they each encode around 12 000 genes, encoding diverse and complex functions likely acquired through their evolutionary history, combining genes from exosymbiont as well as from algal endosymbionts and horizontal gene transfer events (Vardi et al., 2009; Tirichine et al., 2017). Re-sequencing of seven additional *T. pseudonana* strains (Koester et al., 2018) and ten *P. tricornutum* accessions (Rastogi et al., 2020) revealed polymorphisms, providing evidence for genome variation, evolution and adaptation (Falciatore et al., 2020).

More recent studies have extended our knowledge of the complexity of diatom genomes: *Fistulifera solaris* (Tanaka et al., 2015) with an allodiploid genome structure, the cold-adapted pennate *Fragilariopsis cylindrus* (Mock et al., 2017) with a highly heterozygous genome showing allele-specific expression in response to environmental stresses, *Pseudo-nitzschia multiseries*, *Cyclotella cryptica* (Traller et al., 2016), *Pseudo-nitzschia multisistriata* (Basu et al., 2017), *Seminavis robusta* (Osuna-Cruz et al., 2020), and *Thalassiosira oceanica* (Lommer et al., 2012) with more than twice the number of protein-coding genes than *T. pseudonana* and *P. tricornutum*, suggesting the traditional diatom models are underestimating gene diversity (Osuna-Cruz et al., 2020) (Table 1.1).

Table 1.1. Genetic features of established and emerging diatom model species. Table adapted from Falciatore et al. (2020).
<table>
<thead>
<tr>
<th>Species</th>
<th>Strain</th>
<th>Ploidy</th>
<th>Genome size (Mbp)</th>
<th>Number of genes</th>
<th>Average gene size (bp)</th>
<th>Introns per gene</th>
<th>Plastid genome size (kbp)</th>
<th>Mitochondria genome size (kbp)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Thalassiosira pseudonana</em></td>
<td>CCMP 1335</td>
<td>diploid</td>
<td>32.1</td>
<td>11776</td>
<td>1741</td>
<td>1.54</td>
<td>129</td>
<td>43.8</td>
<td>(Armbrust et al., 2004)</td>
</tr>
<tr>
<td><em>Thalassiosira oceanica</em></td>
<td>CCMP 1005</td>
<td>diploid</td>
<td>81.6</td>
<td>34642</td>
<td>1255</td>
<td>1.29</td>
<td>141</td>
<td>35.3</td>
<td>(Lommer et al., 2012)</td>
</tr>
<tr>
<td><em>Cyclotella cryptica</em></td>
<td>CCMP 332</td>
<td>diploid</td>
<td>161.7</td>
<td>21121</td>
<td>1471</td>
<td>1.18</td>
<td>129</td>
<td>58</td>
<td>(Traller et al., 2016)</td>
</tr>
<tr>
<td><em>Phaeodactylum tricornutum</em></td>
<td>CCAP 1055/1</td>
<td>diploid</td>
<td>27.4</td>
<td>12233</td>
<td>1624</td>
<td>1.1</td>
<td>117</td>
<td>77.3</td>
<td>(Bowler et al., 2008; Rastogi et al., 2018)</td>
</tr>
<tr>
<td><em>Fragilariopsis cylindrus</em></td>
<td>CCMP 1102</td>
<td>diploid (triploidy proposed)</td>
<td>61.1</td>
<td>21066</td>
<td>1575</td>
<td>1.38</td>
<td></td>
<td></td>
<td>(Mock et al., 2017)</td>
</tr>
<tr>
<td><em>Pseudo-nitzschia multiseries</em></td>
<td>CLN-47</td>
<td>diploid</td>
<td>218.7</td>
<td>19703</td>
<td>1522</td>
<td>0.87</td>
<td>111.5</td>
<td>46.3</td>
<td>(Basu et al., 2017)</td>
</tr>
<tr>
<td><em>Pseudo-nitzschia multistriata</em></td>
<td>B856</td>
<td>diploid</td>
<td>59</td>
<td>12008</td>
<td>2205</td>
<td></td>
<td></td>
<td></td>
<td>(Tanaka et al., 2015)</td>
</tr>
<tr>
<td><em>Fistulifera solaris</em></td>
<td>JPPC DA0580</td>
<td>allodiploid</td>
<td>24.9</td>
<td>11448</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Tanaka et al., 2015)</td>
</tr>
<tr>
<td><em>Seminavis robusta</em></td>
<td>D6</td>
<td>diploid</td>
<td>151</td>
<td>36254</td>
<td>1801</td>
<td>2.33</td>
<td>150</td>
<td>44</td>
<td>(Osuna-Cruz et al., 2020)</td>
</tr>
</tbody>
</table>
Although the number of diatom genomes available has increased in recent years, this is still limited to only about nine diatoms. So new projects have emerged to increase the possibility to harness diatom complex and unique biology. An example is the “100 Diatom Genomes Project” from JGI (https://jgi.doe.gov/csp-2021-100-diatom-genomes/), which is sequencing one hundred diatom genomes, a science incubator with the aim of providing unique insights into diatom ecological and evolutionary role and of advancing in diatom-based biotechnology and synthetic biology.

In parallel with most diatom genomes, extensive RNA-seq datasets and other omics data were obtained under different growth conditions. The most comprehensive sets of conditions have been profiled for *P. tricornutum* and *T. pseudonana* (Falciatore et al., 2020). Transcriptomes for these species include responses to environmental factors such as light, diurnal cycle, nutrient availability, chemical pollutants and biotic interactions, or physiological features such as growth phase. In other species, transcriptomic datasets were mainly used to answer to specific question related to the species of interest, such as sex determination in *P. multistriata* (Russo et al., 2018a).

In addition to genomes and transcriptomes, proteomics and metabolomics data are available for some diatoms. For example, proteomic profile of *T. pseudonana* (Chen et al., 2018) and *Skeletonema marinoi* (Thangaraj et al., 2020) in different nutrient and temperature conditions highlighted their common or specific responses to different external factors. In the same way, lipid diversity and metabolomic adaptations to N limitation were investigated in thirteen diatom species grown in different N concentrations (Bromke et al., 2015), and metabolomics exploration of the sexual phase was performed in *P. multistriata* (Fiorini et al., 2020).

The availability of many omics datasets opens the possibility to integrate independent information. It has greatly facilitated genome annotations and so gene discovery in many species, to hypothesise the function of unknown genes based on expression patterns, or to actively search for gene families with a specific response. However, very often these data are fragmented and difficult to find and combine. Furthermore, the evolution of sequencing and analysis techniques produces different outputs often difficult to compare. In this context, many online tools have been made available, which can help this kind of research.

To date, different online tools can help to retrieve and compare all these recent data. Marine microalgal genomes can be found in several databases, such as the JGI (Joint
Genome Institute Genome Portal - [https://genome.jgi.doe.gov/portal/](https://genome.jgi.doe.gov/portal/), with genomic data that can be searched, downloaded and explored with several analytical tools (Grigoriev et al., 2012), or Ensembl Protists ([https://protists.ensembl.org/index.html](https://protists.ensembl.org/index.html)), a section of the Ensembl genome browser originally developed for vertebrates by the European Bioinformatics Institute (EBI) and the Wellcome Trust Sanger Institute, which is used for the annotation, analysis and display of genomes (Kersey et al., 2018).

DiatomCyc ([http://www.diatomcyc.org](http://www.diatomcyc.org)) is a comprehensive database on the model pennate diatom *P. tricornutum*, providing information about metabolic pathways, reactions, compounds, proteins and genes (Fabris et al., 2012). Diatom Portal v.1 ([http://networks.systemsbiology.net/diatom-portal/](http://networks.systemsbiology.net/diatom-portal/)) integrates analyses of all publicly available microarray data for the diatoms *T. pseudonana* and *P. tricornutum*, including shared expression patterns, gene functions, and *cis*-regulatory DNA sequence motifs in each species that are statistically coordinated over many experiments (Ashworth et al., 2016). ASAFind ([https://rocaplab.ocean.washington.edu/tools/asafind/](https://rocaplab.ocean.washington.edu/tools/asafind/)) identifies nuclear-encoded plastid proteins in algae with secondary plastids of the red lineage based on the identification of conserved transit peptides (Gruber et al., 2015). Made public more recently, PLAZA diatoms 1.0 ([https://bioinformatics.psb.ugent.be/plaza/versions/plaza_diatoms_01/](https://bioinformatics.psb.ugent.be/plaza/versions/plaza_diatoms_01/)) is an access point for comparative and functional genomics centralising genomic data produced by different genome sequencing initiatives. It integrates sequence data and comparative genomics methods and provides an online platform to perform evolutionary analyses and data mining (Osuna-Cruz et al., 2020). As regard proteomic data, UniProt ([https://www.uniprot.org/](https://www.uniprot.org/)) provides a comprehensive, high-quality and freely accessible resource of protein sequence and functional information, including proteomic data (The UniProt Consortium et al., 2021).

Another important resource is represented by the Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP - [https://www.imicrobe.us/#/search/mmetsp](https://www.imicrobe.us/#/search/mmetsp)), funded by the Gordon and Betty Moore Foundation, integrate microbial eukaryotes into marine ecology by creating 678 assembled, functionally annotated, and publicly available transcriptomes (Keeling et al., 2014). These transcriptomes largely come from the 411 more abundant and ecologically significant microbial eukaryotes in the oceans, 92 of which are diatoms, cultured in different environmental conditions to cover as much as possible nature variability. The
choice of species, strain, and physiological condition was based on different parameters considering phylogeny, physiological impact, environmental and ecological importance of species (Keeling et al., 2014).

The most recent DiatOmicBase (https://www.diatomicsbase.bio.ens.psl.eu/) is a genome portal to perform research on P. tricornutum, gathering comprehensive omics resources to ease the exploration of dispersed public datasets. Here it is possible to find a genome browser, genes pages with models and co-expression networks, and a transcriptomic module. This new tool has been recently expanded with T. pseudonana and P. multistriata. DiatoOmicBase is also adding information coming from proteomic and metabolomic studies, which are often difficult to retrieve due to the lack of unified databases.

1.2.2. Meta-omics revolution

In the last years, the omics disciplines have been joined by the more challenging meta-omics ones. Sequencing of DNA and RNA (i.e. cDNA) from environmental samples without any culturing step are defined as metagenomics and metatranscriptomics respectively, and they allow researchers to investigate how microorganisms distribution is structured, how they interact among them, and how they adapt to their environments (Handelsman et al., 1998; Gill et al., 2006). Due to the recent advances in environmental nucleic acid sequencing and the drop of its cost, many samples were processed in the last years, so that now the main challenge is often represented by the downstream computational analysis.

A very interesting resource is given by TARA Oceans (https://tara-oceans.mio.osupytheas.fr/ocean-gene-atlas/), a circumglobal expedition in which eukaryotic diversity was assessed from 334 size-fractionated photic-zone plankton communities collected across tropical and temperate oceans (de Vargas et al., 2015). It systematically collected more than 35 000 ocean water and plankton samples as well as environmental data at 210 stations, in a period of 3 years (Fig. 1.8). Nucleic acids were extracted from samples and subjected to high-throughput sequencing to generate metabarcoding (metaB), metagenomic (metaG) and metatranscriptomic (metaT) datasets as well as to yield single-cell genomes (Sunagawa et al., 2020). These analyses were coupled with high-throughput imaging tools, which captured abundance and morphological features of plankton across size fractions.
Recently, new release of the “Ocean Gene Atlas” (OGA2 - https://tara-oceans.mio.osupytheas.fr/ocean-gene-atlas/) includes also 58 metagenomes from tropical and subtropical deep oceans to generate the Malaspina Gene Database (Vernette et al., 2022; Acinas et al., 2021; Duarte, 2015).

Figure 1.8. The map shows the cruise track of TARA Oceans as she sailed the world from September 2009 to December 2013 and the location of 210 stations, which were chosen to cross and to sample as many biogeographic provinces and environmental features as possible (sea surface temperature shown as a colour gradient). Overall, more than 35 000 samples of seawater and plankton were collected and archived in partner laboratories. Image adapted from Sunagawa et al. (2020).

The TARA Oceans dataset has been already exploited to study diatoms. Previous studies described the taxonomic abundance and diversity at global scale, highlighting the importance of nanodiatoms (Leblanc et al., 2018). More recently, Busseni et al. (2019) explored global diatom diversity at different scales from taxonomic to functional, focusing on N transporters such as NO$_3^-$ high-affinity transporters (NRT2) and ammonium (NH$_4^+$) transporters (AMT), while Vincent and Bowler (2020) explored diatom biotic interactions as drivers of diatom evolution and adaptation in the modern ocean.

Recently, in the context of the meta-omics data revolution, a new concept of Planetary Biology is emerging, combining direct investigation of diverse planetary ecosystems from field studies with controlled laboratory research on experimental model systems, with the aim of unravelling genetic and environmental influences in biomes and
ecosystems, of understanding biological processes and of improving planetary health (https://www.embl.org/about/programme/research-plans/planetary-biology/).

1.2.3. Genetic tools for diatoms

A genomic sequence can be a tool to reconstruct pathways, and comparative genomics can help to identify the genetic basis of particular traits and to characterise genes responsible for a given phenotype. This can be possible through genetic engineering, a promising approach to improve the metabolic potential of diatoms and obtain further insights into their metabolism and physiology (Serif et al., 2018). The accumulation of genomic and transcriptomic data associated with the development of cheap and easy-to-use genome editing technologies are facilitating the continuous and rapid progress in diatom biology.

The breakthrough that made reverse genetics in diatoms possible was the biolistic delivery of transgenes to *C. cryptica* and *Navicula saprophila* cells (Dunahay and Jarvis, 1996) (Fig 1.9A and Table 1.2). This technique consists on the adhesion of exogenous DNA to gold or tungsten microspheres surface, which are "shot" at a particular pressure on target cells to be transformed (Falciatore et al., 1999; Poulsen et al., 2006; Zaslavskaia et al., 2001; Dunahay and Jarvis, 1996). Biolistic has been the most used transformation technique, in particular on *P. tricornutum* and *T. pseudonana*, where it was performed with modifications which gave reproducible results even though with low transformation efficiencies ($10^{-6}$ to $10^{-8}$ transformants per µg of DNA) (Falciatore et al., 1999, 2020).

DNA can be delivered to *P. tricornutum* also via electroporation (Niu et al., 2012) (Fig 1.9B and Table 1.2). This technique uses electric pulses to create transient pores in the cell membrane, which promote the delivery of genetic material into cells.

The major issue of this method and the biolistic one is the random integration of the transgene in the genome, often with multiple integration events. Chromosomal position effects and variable transgene copy numbers cause significant variability in transgene expression levels in different clones (Falciatore et al., 2020). Moreover, biolistic transformation acts with a physical force that can produce double-strand breaks subsequently repaired by non-homologous end-joining, affecting genome integrity (Falciatore et al., 2020).

To avoid random integrations in the genome, recently a new transformation method has been developed, called bacterial conjugation (Karas et al., 2015) (Fig 1.9C and Table 1.2).
It is based on a plasmid containing a yeast-derived sequence, which can be delivered to *P. tricornutum* and *T. pseudonana* through bacterial conjugation using *Escherichia coli* and which enables stable episome replication in these diatoms in the presence of selective antibiotic pressure, showing that episomes are maintained as closed circles at copy number equivalent to native chromosomes. This transformation strategy is more efficient than previously described methods (10⁻⁴ transformants per µg of DNA) but less stable, as episomes are eventually lost in the absence of selective pressure (Diner et al., 2016; Karas et al., 2015; Falciatore et al., 2020). Bacterial conjugation can be particularly useful to achieve transient expression, in case of potentially deleterious transgenes.

These transformation methods can be used for various purposes, as the inactivation or overexpression of genes, or the localisation of a gene product in live cells through fusion with fluorochromes, such as the Yellow Fluorescent Protein (YFP) and Green Fluorescent Protein (GFP).

In particular, many vectors have been designed for protein tagging and promoter-reporter/target transgenes to study gene functioning. In *P. tricornutum* and *T. pseudonana*, *Lhcf* (Light-Harvesting Complexes) promoters are most used to drive strong expression of target genes (Poulsen et al., 2006; Falciatore et al., 1999; Apt et al., 1996), while the Nitrate Reductase *NR* promoter, originally developed for the diatom *Cylindrotheca fusiformis* (Poulsen and Kröger, 2005a), allows gene expression to be induced by shifting the transgenic cells from nitrate (NO₃⁻)-deficient to NO₃⁻-supplemented medium (Hempel et al., 2009; Poulsen et al., 2006) (Table 1.2). Moreover, the histone *H4* promoter has been successfully used in *P. tricornutum* and *P. multistriata*.
to obtain a mild but constitutive and stable overexpression of the gene of interest (Siaut et al., 2007; Sabatino et al., 2015). Among the most frequently used reported genes, are fluorescent tags for subcellular localisation of proteins like \textit{eGFP} (Poulsen et al., 2006; Zaslavskia et al., 2001) and \textit{eYFP} (Karas et al., 2015; Siaut et al., 2007). The most common selectable markers for antibiotic resistance are \textit{Sh-ble} for zeocin and phleomycin resistance (Apt et al., 1996), \textit{Nat} for nourseothricin, \textit{Sat} for streptothricin and \textit{NptII} for neomycin (Zaslavskia et al., 2001) (Table 1.2).

Furthermore, it is possible, by exploiting the strategy of RNA interference (RNAi), to obtain stably knock-down, or silenced clones (De Riso et al., 2009), although little is known about how gene silencing works in diatoms (Rogato et al., 2014) (Table 1.2).

**Table 1.2.** Technologies and genome-enabled tools available for diatoms. Adapted from Falciatore et al. (2020).

<table>
<thead>
<tr>
<th>Technologies</th>
<th>Available tools</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transformation</td>
<td>Biolistic</td>
<td>(Dunahay and Jarvis, 1996)</td>
</tr>
<tr>
<td></td>
<td>Electroporation</td>
<td>(Niu et al., 2012)</td>
</tr>
<tr>
<td></td>
<td>Conjugation</td>
<td>(Karas et al., 2015)</td>
</tr>
<tr>
<td>Gene silencing</td>
<td>Inverted repeats and antisense fragments</td>
<td>(De Riso et al., 2009)</td>
</tr>
<tr>
<td>Genome editing</td>
<td>Meganucleases</td>
<td>(Daboussi et al., 2014)</td>
</tr>
<tr>
<td></td>
<td>TALEN</td>
<td>(Daboussi et al., 2014)</td>
</tr>
<tr>
<td></td>
<td>CRISPR/Cas9</td>
<td>(Nymark et al., 2016)</td>
</tr>
<tr>
<td></td>
<td>Delivered as RNPs</td>
<td>(Serif et al., 2018)</td>
</tr>
<tr>
<td>Most used promoters</td>
<td>\textit{Lhcf} (light sensitive – strong)</td>
<td>(Apt et al., 1996)</td>
</tr>
<tr>
<td></td>
<td>\textit{NR} (nitrate inducible – strong)</td>
<td>(Hempel et al., 2009)</td>
</tr>
<tr>
<td></td>
<td>\textit{H4} (constitutive – mild)</td>
<td>(Siaut et al., 2007)</td>
</tr>
<tr>
<td>Antibiotic resistance</td>
<td>\textit{Sh-ble} (phleomycin, zeocin)</td>
<td>(Apt et al., 1996)</td>
</tr>
<tr>
<td></td>
<td>\textit{Nat} (nourseothricin)</td>
<td>(Zaslavskia et al., 2001)</td>
</tr>
<tr>
<td></td>
<td>\textit{Sat} (streptothricin)</td>
<td>(Zaslavskia et al., 2001)</td>
</tr>
<tr>
<td></td>
<td>\textit{NptII} (neomycin)</td>
<td>(Zaslavskia et al., 2001)</td>
</tr>
</tbody>
</table>

### 1.2.4. Genome editing

Genome editing approaches allow the direct modification of one or more copies of a gene within a cell using engineered nucleases (Kroth et al., 2018). One of the first systems applied in diatoms was based on meganucleases (Daboussi et al., 2014), which are particular endonucleases able to bind specific target sequences on genome and then
replace, eliminate or modify those sequences. However, meganucleases production is
time-consuming and not very flexible (Kroth et al., 2018). So, more recently, the search
for new techniques for genome editing allowed the development of more modern systems,
such as TALEN (Transcription Activator-Like Effector Nucleases) and CRISPR
(Clustered Regulatory Interspaced Short Palindromic Repeats) / Cas9, allowing the
generation of knock-out strains (Nymark et al., 2016; Daboussi et al., 2014).

All these approaches, from the meganucleases to the CRISPR/Cas9 system, have in
common that they may induce a Double-Strand Break (DSB). DSB damage can then be
repaired either through the Non-Homologous End-Joining (NHEJ) mechanism, which
could induce interruption site mutations such as insertions/deletions (INDELs) or single
nucleotide replacements, or more rarely through the Homology Directed Repair
mechanism (HDR), which instead can introduce insertions when a donor DNA strand is
made available in the cell. In many organisms, HDR has been used to introduce precise
DNA mutations, to insert specific sequences in a target locus, or to achieve targeted gene
replacements (Rath et al., 2015).

TALE nucleases are chimeric proteins made by fusing a TALE DNA-binding domain
designed to recognise and bind a specific sequence with a non-specific catalytic head, the
FokI endonuclease (Bitinaite et al., 1998). Except for two polymorphic residues known
as Repeat Variable Di-residues (RVDs), which are found at positions 12 and 13 and which
are crucial for the specific recognition of a target nucleotide, the binding domain is made
of up to 14–24 repeat units of 33–35 essentially identical amino acids. Based on this
characteristic, the DNA-binding domain can be modified to detect almost any sequence
(Kroth et al., 2018). As FokI functions only as a dimer, two independent monomeric
TALENs are required to generate a DSB (Bitinaite et al., 1998). TALEN system has been
recently used in diatoms, knocking out a key gene for storage carbohydrate synthesis in
P. tricornutum (Daboussi et al., 2014) and the nitrate reductase gene (McCarthy et al.,
2017).

CRISPRs have been discovered in the genomes of bacteria and archaea and consists of
repeated sequences interrupted by sequences previously absorbed by genomic viruses.
These sequences provide adaptive immunity through the CRISPR associated protein,
called Cas (CRISPR Associated), which acts as an RNA-driven endonuclease and is
capable of destroying the same type of invasive DNA if it is met again (Doudna and
Charpentier, 2014). To date, three CRISPR/Cas systems have been classified based on
the presence of a specific Cas protein subtype: Type I, derived from *Escherichia coli* and *Pseudomonas aeruginosa*, targets DNA sequences with the help of the metal-dependent endonuclease activity of Cas3 protein; Type II derived from *Streptococcus thermophilus* and *Streptococcus pyogenes*, and acting through nuclease Cas1, Cas2 or Cas9 (formerly Csn1); and Type III derived from *Staphylococcus epidermi*, *Lactococcus lactis* and *Pyrococcus furiosus* (Kumar and Jain, 2015). Among these, the type II system derived from *Streptococcus pyogenes* is the most commonly used for genetic engineering applications.

So, the CRISPR / Cas9 system, integrated into the natural defence mechanism of bacteria, is the new technology based on the use of the Cas9 protein, an endonuclease that is able to cut a specific DNA target through a DSB, once a single-strand guide RNA (gRNA) is provided, with a sequence complementary to the target gene (Fig. 1.10). The gRNA contains a targeting sequence (crRNA sequence) homologous to the genomic region to be modified, and a Cas9 nuclease-recruiting sequence (tracrRNA). The binding specificity is based on the gRNA and a 3-nucleotide downstream sequence called the Protospacer Adjacent Motif (PAM - Fig. 1.10) (Kroth et al., 2018).

Figure 1.10. Schematic representation of the CRISPR/Cas9 system.
The CRISPR / Cas9 system is a successful tool to knock out genes, and also allows to work with gene families with functional redundancy, through the possibility of knocking out multiple genes within the same family in order to generate a phenotype (Sharma et al., 2021a). Despite the success of the CRISPR / Cas9 system, this technique can induce the occurrence of unwanted off-target mutations (Kroth et al., 2018). Although there are tools for predicting potential off-targets, they are not always accurate because the enzyme may ignore some anticipated off-target sites while randomly introducing additional DSBs or cuts in unpredicted sites (Anderson et al., 2015). Using inducible and/or repressible promoters that enable temporally controlled nuclease expression is one technique to reduce the amount of time the genome is exposed to nuclease activity with the goal of reducing non-specific cuts.

Among other strategies to minimise the exposure of the genome to nuclease activity, with the aim of reducing non-specific cuts, there is the creation of the SpCas9-HF1, a variation of the *S. pyogenes* Cas9 (SpCas9-High Fidelity 1) (Kleinstiver et al., 2016). Due to a mutation in the nuclease domain, this variant can reduce non-specific interactions with the genome and is able to recognise highly specific gRNA without mismatches and allelic variants on target sequences (Kleinstiver et al., 2016). Another way is the bacterial conjugation, for which the Cas9 encoding gene is transferred to cell via an episomial vector carrying resistance gene. The episomal vector can be subsequently eliminated from the diatom cells by removal of the selective pressure, resulting in transient Cas9 expression and non-transgenic mutant lines (Sharma et al., 2018; Russo et al., 2018b). Depending on country legislation, such lines might be considered as non-Genetically Modified Organisms (GMO) (Sharma et al., 2018).

Finally, the most recent system exported on diatoms is proteolitics (Serif et al., 2018), which involves the use of ribonucleoprotein complexes (RNP), formed by RNA guides and the nuclease Cas9. The two most important advantages of this protocol are the reduction of off-targets, avoiding random integration of exogenous DNA within the host genome and long-term expression of the nuclease, and the selection based on the knockout of an endogenous gene, that allows an antibiotic-free selection (Serif et al., 2018). More detailed description of CRISPR/Cas9 system and proteolistic protocol is given in the paragraph 3.1 “Introduction” in Chapter 3.
1.2.5. The model diatom *Phaeodactylum tricornutum*

*P. tricornutum* is a marine pennate diatom firstly described by Bohlin in 1897 in samples collected in Plymouth and Baltic rock pools. It is poorly silicified and shows several morphotypes, among which fusiform, triradiate, oval, cruciform and round (De Martino et al., 2007) (Fig. 1.11); changes in cell shape can be stimulated by environmental conditions. No substantial evidence for sexual reproduction was recorded (Bowler and Falciatore, 2019).

To date, 10 ecotypes, also called accessions, have been characterised concerning their genotype and features, with a conserved genetic and functional makeup, as a consequence of the limited dispersal of *P. tricornutum* in the open ocean (De Martino et al., 2007; Rastogi et al., 2020, 2018). For example, seven strains are predominantly fusiform, whereas only the strain Pt10 is predominantly triradiate, and Pt9 is predominantly oval and defined as a tropical strain because it appears better acclimated to growth at higher temperatures (De Martino et al., 2007). In particular, the Pt1 strain was used to sequence the genome (Bowler et al., 2008) and recently re-annotated to update the gene models catalogue (Rastogi et al., 2018). Moreover, Pt4 has a particularly reduced photoacclimation capacity reflecting its adaptation to high latitudes, with low light intensities and less drastic diurnal light variations (Bailleul et al., 2010).

Data coming from 80 libraries of RNA-seq generated using different sequencing technologies (among them McCarthy et al., 2017; Smith et al., 2019), two methylome studies (Veluchamy et al., 2013, 2015), five small non-coding RNA libraries (Rogato et al., 2014) and a study on long intergenic non-protein coding RNAs (lincRNAs) (Cruz de Carvalho et al., 2016) are available on public databases (Fig. 1.12). The big availability of omics resources and the relative ease of cultivation in the laboratory, with high growth rates and stability, makes *P. tricornutum* a well-established model species (Bowler and Falciatore, 2019).
Figure 1.11. The pennate diatom *P. tricornutum* showed in its three main morphotypes through light microscope: left, fusiform; top right, triradiate; bottom right, oval. Image adapted from Vardi et al. (2009).

Moreover, several genetic tools are available (Siaut et al., 2007) like genetic transformation (Apt et al., 1996; Falciatore et al., 1999; Karas et al., 2015; Zhang and Hu, 2014) and genome editing techniques (Daboussi et al., 2014; De Riso et al., 2009; Nymark et al., 2016; Serif et al., 2018) (Fig. 1.12). Currently, it represents one of the best and most used models to understand diatom biology, indeed it has been particularly useful to reveal insights into diatom carbon, nitrogen and iron metabolism, as well as cell cycle.
Figure 1.12. Timeline of the major publications and discoveries dealing with the model diatom *P. tricornutum*.

*P. tricornutum* is also used in a large range of biotechnological applications, ranging from the medical and pharmaceutical field (Fabris et al., 2020; Gille et al., 2019) to plastic biodegradation (Moog et al., 2019), displaying potentially promising and sustainable possibilities. In particular, *P. tricornutum* emerged as a potential microalgal energy source, as its lipid storage constitute about 20-30% of its dry cell weight under standard culture conditions. N limitation can induce neutral lipid accumulation in *P. tricornutum*, indicating possible strategies for improving microalgal biodiesel production (Yang et al., 2013).

1.3. Nutrient transporters

1.3.1. To cope with a variable environment

While in the past microbial oceanography was based on the assumption that molecules and organisms are randomly distributed, now there are evidences that physico-chemical parameters are not homogeneously allocated both at macro- and at microscale, but present fast fluctuations, sometimes of one order of magnitude (Stocker, 2012). In this environment, microorganisms can exploit heterogeneity by moving towards favourable conditions or by activating mechanisms to survive in unfavourable ones. Moreover, movements at the larger scales and the seasonal cycle induce further changes in nutrient availability that microorganisms need to adapt to Rogato et al. (2015).

Variations in seawater conditions directly affect phytoplankton, whose growth and distribution is strongly determined by light and nutrient conditions. Primary producers distribution influences that of primary consumers and so on, thus variable conditions of the ocean dynamically shape the structure of entire communities and ecosystems.

One of the most influencing abiotic variables is light, an essential resource for diatom growth which provides a significant amount of information about the surrounding environment, through its intensity, orientation, wavelength and timing. For example, the light intensity and the relative ratio of different wavelengths can be affected by the filtration activity of a layer of light-absorbing elements, while the light timing can change because of daily light/dark cycle together with the periodic variation in day length across
seasons and latitudes. Then, light is absorbed by water itself, mostly in the red part of the spectrum, so the blue and green bands penetrate the water column more deeply (Jaubert et al., 2022).

Light features are sensed by diatoms, which then need to respond to those particular conditions. Light sensing relies on pigment-bound proteins called photoreceptors, that absorb light of specific wavelengths and activate specific signalling cascades (Duanmu et al., 2017). There are also other photosynthetic pigments which collectively absorb the larger light fraction, and photosynthetic activities in plastids which regulate the light responses through retrograde pathways from plastids to the nucleus (Jaubert et al., 2022).

Another important abiotic factor influencing diatom growth and distribution is the availability of macro- and micronutrients. Macronutrients are essential elements required in relatively high concentrations to support cellular growth and metabolism. They include the most abundant elements in cells, such as nitrogen (N), phosphorus (P), sulphur (S) and carbon (C). In addition, most diatoms have a silicon (Si) requirement in order to synthesise their cell wall, or frustule (Smith and Allen, 2022). On the other hand, micronutrients are typically required in quantities at least three orders of magnitude less than macronutrients, and their cellular stoichiometry is much more variable: they include metals such as iron (Fe) and zinc (Zn), but also manganese (Mn), copper (Cu) and cobalt (Co), as well as vitamins like cobalamin (B$_{12}$), thiamine (B$_{1}$), and biotin (B$_{7}$) (Jaubert et al., 2022; Tréguer et al., 2018; Coale et al., 2022).

Both planktonic and benthonic diatoms experience strong fluctuations in nutrient concentrations due to currents, water mixing, tides and seasons (Fig. 1.13). In this context, diatoms have the capacity to adjust their physiology thanks to different strategies for acquisition, storage, and metabolic usage, which help them to cope with these variations.
1.3.2. Kinetics of nutrient transporters

To cope with changes in environmental conditions and in particular nutrient availability (Fig. 1.13), phytoplankton have developed different strategies, by altering their chemical composition in response to environmental variability. This acclimation capabilities allow organisms to maintain their growth rate in a changing and often adverse environment as well as to avoid possible damages due to an overexposure to the resources (Bonachela et al., 2011).

A particular feature of phytoplankton acclimation is its ability to adjust the uptake machinery in response to changes in nutrient availability, but also to internal physiological conditions (Bonachela et al., 2011; Rogato et al., 2015). In this context, proteins as transporters and channels play a key role in regulating acclimation processes,
and so in determining the ecological niche and, ultimately, overall community composition.

Nutrient uptake can be generally formulated by Michaelis–Menten kinetics:

\[
\text{Uptake-rate} = \text{Uptake-rate}_{\text{max}} [\text{Nut}] / (K_{\text{Nut}} + [\text{Nut}])
\]

with \([\text{Nut}]\) the medium concentration of the specific Nutrient and \(K_{\text{Nut}}\) the half-saturation constant related to the affinity of the cell for a specific nutrient. However, treating \(\text{Uptake-rate}_{\text{max}}\) as a constant and considering the uptake response as that of an enzyme, prevents the classic Michaelis–Menten formulation from considering acclimation, which can be identified, in this case, as the ability of the cell to adjust its kinetic parameters in response to environmental changes. So Bonachela et al. (2011) added a modulating term dependent on the internal stock of the considered nutrient:

\[
\text{Uptake-rate} \propto \text{Growth-rate} = \text{Growth-rate}_{\text{max}} (1 - Q_{\text{max}} / Q)
\]

where the intracellular C-based Quota, \(Q = [\text{Nut}]/[C]\), is a proxy for the need (or absence thereof) to synthesise new transporters (Rogato et al., 2015; Bonachela et al., 2011). Merging the last two concepts, the dependence of growth from uptake rate would be a function of external and internal concentrations with the maximum assimilation rate \((\text{Uptake}_{\text{max}})\) modulated by the internal nutrient quota:

\[
\text{Uptake} \propto (1 - Q/Q_{\text{max}}) / (1 - Q /Q_{\text{max}} + \text{shape function}) \ast [\text{Nut}] / (K_{\text{Nut}} + [\text{Nut}]),
\]

For a given value of C, the quota is low if the internal nutrient concentration is low for the current size or biomass of the cell, and \textit{vice versa}. Thus, there will be upregulation if the cell needs to incorporate nutrient \((Q \text{ close to its minimum})\), and downregulation when the performance of the intake of nutrient is sufficiently high so that \(Q\) is large and the cell can reduce the number of transporters, in order to also decrease maintenance and biosynthetic costs (Bonachela et al., 2011).

Also luxury uptake, which means an uptake rate larger than what would be required to sustain growth, is a factor that needs to be considered as it influences transporters regulation, and in particular with an opposite correlation. In fact, from this model, it follows that the cell upregulates transporters synthesis to exploit the excess of internal resources, however, actually, the cell downregulates the production of transporter for energetic purposes, saving energy when nutrient availability is low (Bonachela et al., 2011).
More dynamic variables were added or still needs to be added to this assumption, to better describe nutrient transport: for example time, the absorbing area compared to the total cell area, the transporters affinity, the activation or repression (positive or negative feedback) of gene expression which determine the number of available transport sites, and transport contribution from multiple compartments, not only from external environment but also from storage organelles. Moreover, a cross-talk among different signalling pathways involved in different nutrients metabolisms has to be considered for a correct decoding of the nutritional status of cells, together with the plasticity in the uptake rate which rely on multivariate control systems acting on the key components, the membrane transporters (Rogato et al., 2015).

This highlights nutrient transport system as a flexible, dynamic process, able to respond to multiple changes in biogeochemical fluxes (Bonachela et al., 2011; Van de Waal and Litchman, 2020) but also in cell physiological conditions. In fact, the regulation of nutrient uptake may be involved in the general control of whole cell cycle and life strategies, considering the need to distribute metabolic energy and efforts to different competing activities such as photosynthesis, metabolite production or sexual reproduction. Interestingly, in the sexually reproducing diatom *P. multistriata*, growth can be arrested by specific signals during mating even in nutrient replete conditions (Scalco et al., 2016; Annunziata et al., 2022).

1.3.3. Nutrients and their specific transport

As previously mentioned, key nutrients, beside C, are P, N and, specifically for diatoms, Si. Lack of these nutrients prevents phytoplankton growth, for this reason they are defined as limiting or co-limiting nutrients (Rogato et al., 2015).

Diatoms have approximately 1000 transporters per genome identified by the Transporter Classification Database (TCDB) together with Phycocosm (https://phycocosm.jgi.doe.gov/mycocosm/annotations/browser/tcdb/summary) (Smith and Allen, 2022) (Fig. 1.14). Of these, the most abundant and diverse types of macronutrient transporters identified were for N, including different forms such as nitrate (NO$_3^-$), nitrite (NO$_2^-$), ammonium (NH$_4^+$), urea, and amino acids (116 combined average per genome), for phosphate (PO$_4^{3-}$) (17 average), followed by putative sulphate transporters (11 average) and Si transporters (6 average) (Smith and Allen, 2022) (Fig. 1.14). These data included transporters targeted both to the plasma membrane and
intracellular compartments, since there is not always localisation information. Below, an overview of key membrane nutrient and ion transport processes.

![Figure 1.14. Major classes and abundance of macronutrient transporters from diatom genomes. Genes were identified from the literature and transporter annotations from the Transporter Classification Database (TCDB) obtained from Phycocosm (a comparative algal genomics resource) for ten diatoms (Smith and Allen, 2022). Box and whisker plot shows distribution of data into quartiles, with the central line indicating the exclusive median, and x indicates the mean. Data labels show the mean value. Image adapted from Smith and Allen (2022).](image_url)

1.3.3.1. Phosphorus (P)

P is an essential element to life, providing the PO$_4^{3-}$-ester backbone of DNA and RNA and being crucial in the transmission of chemical energy through the ATP molecule (Paytan and McLaughlin, 2007). Diatoms can readily access P as inorganic PO$_4^{3-}$ ($P_i$). However, $P_i$ is often found at lower concentrations than Dissolved Organic Phosphorus (DOP), with large variations in relation to several factors, such as depth, and being an abiotic factor limiting productivity in the ocean. In particular, in the open ocean, $P_i$ is rapidly utilised by the phytoplanktonic and bacterial organisms within the euphotic zone.
(until 80-200 meters), where superficial P\textsubscript{i} concentration ranges from 0.2 nM of Sargasso Sea to 1-3 μM of eastern margins of the Atlantic and Pacific Oceans (Paytan and McLaughlin, 2007).

To cope with generally low P\textsubscript{i} concentrations, diatom P\textsubscript{i} demand can be decreased by substituting plasma membrane phospholipids with non-phosphorus forms, or by reducing C and N assimilation and growth (Feng et al., 2015; Brownlee et al., 2022). Other strategies include the secretion of alkaline phosphatases that can increase P\textsubscript{i} availability from DOP (Dell’Aquila et al., 2020; Lin et al., 2013). In \textit{P. tricornutum}, PtPhos1, 2, 3, and 8 are secreted phosphatases that catalyse the hydrolysis of a phosphoric acid monoester/diester, thereby providing P\textsubscript{i} for further uses (Dell’Aquila et al., 2020). In particular, PtPhos1 and PtPhos2 were first identified by analysing the medium for secreted proteins (Lin et al., 2013).

Diatoms own also a big number of P\textsubscript{i} transporters, significantly induced in P\textsubscript{i} limitation. Several \textit{P. tricornutum} putative sodium (Na\textsuperscript{+})/P\textsubscript{i} symporters (PtNap\textsubscript{i}1–6), a P\textsubscript{i} permease (PtPho4) and two putative protons (H\textsuperscript{+})/P\textsubscript{i} transporters (PtHp\textsubscript{i}1 and 2) showed increased expression during P limitation in \textit{P. tricornutum} (Dell’Aquila et al., 2020) and a subcellular localisation either to the plasma membrane or to an endoplasmic reticulum (ER)-like endomembrane (Figs. 1.14 and 1.15). To note that Na\textsuperscript{+}/P\textsubscript{i} symporters, P\textsubscript{i} permeases and H\textsuperscript{+}/P\textsubscript{i} transporters have different substrates, both anions (HPO\textsubscript{4}\textsuperscript{2–} and H\textsubscript{2}PO\textsubscript{4}–) and cations (H\textsuperscript{+} and Na\textsuperscript{+}), which could have an impact on regulation of P\textsubscript{i} uptake and/or maintenance of intracellular ion homeostasis in different taxa (Smith and Allen, 2022).

The ability of diatoms to adapt their metabolism to P\textsubscript{i} limitation needs mechanisms for sensing and responding to exogenous P\textsubscript{i} levels, but also to store cellular P\textsubscript{i} to maintain PO\textsubscript{4}\textsuperscript{3–} homeostasis. The Vacuolar Transport Chaperones (VTCs), which consist of at least four subunits, are known to be involved in synthesis of the stored polyphosphate (polyP). They are SPX domain-containing proteins, important players in regulating responses to P\textsubscript{i} limitation (Brownlee et al., 2022), and include a vacuolar transport chaperone (VTC2) and another one that localises to endomembranes (VTC4) (Dell’Aquila et al., 2020) (Fig. 1.15). Moreover, a homologue of plant vacuolar P\textsubscript{i} transporters (PtVPT1), containing a SPX domain and localised to the vacuolar membrane, with a strong upregulation under P-limiting conditions, could play an important role in P\textsubscript{i} vacuole storage (Dell’Aquila et al., 2020). Interestingly, a gene possessing a SPX domain as the sole functional domain
(named SPX gene) upregulated during Pi limitation in *P. tricornutum*, was knocked out, leading to increases in alkaline phosphatases and Pi transporters and suggesting that it may function as a negative regulator of Pi limitation responses, similarly to plants (Zhang et al., 2021; Brownlee et al., 2022).

1.3.3.2. Silicon (Si)

As previously described, diatoms can accumulate silicic acid (Si(OH)₄) and are capable of creating a Si-based extracellular skeleton, being largely responsible for production of biogenic silica in the global ocean (Sapriel et al., 2009). In *T. pseudonana*, non-saturable Si uptake from high external Si(OH)₄ concentrations (>30 μM) occurs most likely through diffusion, while it is imported with Silicon Transporters (SITs) at lower concentrations (Thamatrakoln and Hildebrand, 2008; Durkin et al., 2016).

Molecular characterisation of the three SITs of *T. pseudonana* indicated that two of them (TpSIT1 and TpSIT2) are localised to the plasma membrane (Fig. 1.15), while TpSIT3 was found in multiple intracellular locations (Shrestha and Hildebrand, 2015). Moreover, the simultaneous RNAi knock-down of *TpSIT1* and *TpSIT2* led to a reduced Si uptake and a lower silica incorporation rate into the frustule (Shrestha and Hildebrand, 2015). In *P. tricornutum*, the three *PtSITs* were shown to be present as a gene cluster, suggesting genome-level regulation of their expression and showing a different gene expression in relation to different morphologies (Brownlee et al., 2022; Sapriel et al., 2009).

Much less is known about the intracellular transporters that mediate Si uptake into the acidic Silica Deposition Vesicle (SDV). An homologue of *LSI2* (Low Silicon Rice 2), proposed to encode for a H⁺/Si(OH)₄ antiporter and regulate Si efflux in plants, was identified in a cluster of genes of *T. pseudonana*, with transcriptional regulation similar to *TpSIT2* (Shrestha et al., 2012; Brownlee et al., 2022) (Fig. 1.15). Diatom LSI2 is a potential candidate for regulating biosilification process in diatoms.

1.3.3.3. Carbon (C)

As regards C, the concentration of Dissolved Inorganic Carbon (DIC) in seawater is around 2 mM, not limiting for photosynthetic organisms, but only a small proportion of around 10 μM is present as CO₂ (Brownlee et al., 2022; Shen et al., 2017). So diatoms need to access the much larger pool of bicarbonate (HCO₃⁻) to maintain the DIC supply for carbon fixation, and they can do it either by direct uptake of HCO₃⁻ or through the action of external Carbonic Anhydrases (eCAs - Fig. 1.15), which catalyse the
interconversion between HCO$_3^-$ and CO$_2$ to maintain an high CO$_2$ concentration at the cell surface (Matsuda et al., 2017; Brownlee et al., 2022). Many other carbonic anhydrases have been found in different subcellular compartments: the two β-type CAs in *P. tricornutum*, PtCA-I and -II, were shown to localise in the periplastidial compartment (Tanaka et al., 2005; Moog et al., 2011); the β-type PtCA-III, -VI and -VII were found in the chloroplast endoplasmic reticulum, which is the part of the endoplasmic reticulum where nuclear-coded proteins that are destined for the chloroplast are synthesised; and the γ-type PtCA-VIII was localised in the mitochondria (Tachibana et al., 2011).

CO$_2$ uptake in diatoms is possible through the high permeability of plasma membrane (Hopkinson et al., 2011). To enhance this capability, *P. tricornutum* possesses five aquaporins, that localise to the plasma membrane, tonoplast and chloroplast endoplasmic reticulum (Matsui et al., 2018), and which work as channel for NH$_3$ and CO$_2$, indicating an important role in facilitating the diffusion of these molecules into and out of diatom cells.

Diatoms can actively take up not only CO$_2$ but also HCO$_3^-$ for photosynthesis, in particular in conditions of low CO$_2$ availability (Shen et al., 2017). The *P. tricornutum* genome encodes seven members of the Solute Carrier SLC4 transporter family and three members of the SLC26 one. One of these encodes PtSLC4–2, which localises to the plasma membrane. Its constitutive overexpression led to high affinity, Na$^+$-dependent HCO$_3^-$ uptake even in high CO$_2$ grown cells, suggesting to play a key role in active HCO$_3^-$ uptake by *P. tricornutum* and other diatoms (Nakajima et al., 2013) (Fig. 1.15).

An interesting and critical aspect of the diatom Carbon Concentrating Mechanism (CCM) is the active transport of HCO$_3^-$ into the chloroplast (Hopkinson et al., 2011). The “chloroplast pump” model enables accumulation of HCO$_3^-$ within or close to the chloroplast, where the activity of carbonic anhydrases can provide a saturating CO$_2$ concentration for fixation by Rubisco (Brownlee et al., 2022; Hopkinson, 2014). However, the identity of the chloroplast HCO$_3^-$ transporters remains still unknown.
Figure 1.15. Compartmentalised model of diatom nutrient transport and storage. Colour gradients indicate putative dissolved organic matter, nutrient, or pH gradients. See text for abbreviations. NO$_3^-$: nitrate, NO$_2^-$: nitrite, NH$_4^+$: ammonium, P$_i$: inorganic phosphate, Si(OH)$_4$: silicic acid, CO$_2$: carbon dioxide, HCO$_3^-$: bicarbonate, Fe$^{2+}$ and Fe$^{3+}$: ferrous and ferric states of iron, Fe$^{3+}$-sid: iron-bounded siderophore, HPO$_4^{2-}$: hydrogen phosphate. NRT2: high-affinity Nitrate Transporter, NPF: low-affinity Nitrate Transporter, AMT: Ammonium Transporter, UT: Urea Transporter, AAT: Amino Acids Transporter, HP$_i$ and NaP$_i$: H$^+$/P$_i$ and Na$^+$/P$_i$ Transporters, SIT: Silicon Transporter, LS12: Low Silicon Rice 2, SLC: Solute Carrier, FBP/FRE: Ferrochrome Binding Protein/Ferric Reductase, ISIP: Iron Starvation-Induced Protein, CLC: Chloride Channel, VPT: Vacuolar Phosphate Transporter, SDV: Silica Deposition Vesicle. Image created with BioRender, following Smith and Allen (2022) and Brownlee et al. (2022).

1.3.3.4. Iron (Fe)

Differently from previously described macronutrients, micronutrients are needed by the microorganisms in very small amounts. Among them, Fe is essential as it is involved in many metabolic reactions: it constitutes important electron transfer proteins involved in the photosynthetic and respiratory pathways, like cytochromes and ferredoxins, but is also present in NO$_3^-$ and NO$_2^-$ reductases and antioxidants enzymes (Street and Paytan, 2005). Moreover, Fe is important for chlorophyll $a$ synthesis. Thus, it is an essential nutrient for planktonic growth, however, in vast areas of the oceans, including most of the Southern Ocean, Fe concentrations are very low, until less than 1 nM (Street and Paytan, 2005).
The ability of oceanic phytoplankton to maintain high growth rates at low Fe concentrations when compared to coastal ones has been well established, particularly in open-ocean regions of the Southern Ocean, equatorial Pacific Ocean and north Pacific Ocean (Marchetti et al., 2006; Armbrust, 2009). These high-nutrient, low-chlorophyll (HNLC) regions are characterised by exceedingly low concentrations of Fe and high concentrations of other essential nutrients, such as N, P and Si.

Diatoms are able to accumulate Fe, which can exist as either Fe$^{2+}$ or Fe$^{3+}$ redox states, from extremely low (sub-nM) concentrations. They have evolved sophisticated mechanisms for Fe acquisition, involving either direct uptake of Fe$^{2+}$ or precisely controlled endocytotic uptake of bound or chelated Fe$^{3+}$ (Brownlee et al., 2022). In fact, the contemporary surface ocean is oxygenated and well mixed, with a mildly alkaline pH (global averages are around 8.1), which makes it an oxidising environment that chemically shifts iron into its ferric state (Fe$^{3+}$) (Gao et al., 2021).

Centric and pennate diatoms differ in their ability to store Fe: for example, the open-ocean centric *Thalassiosira* has permanently modified its photosynthetic apparatus to require less Fe, so compromising their ability to deal with the rapid light fluctuations; while pennate diatoms such as *Phaeodactylum* and *Pseudo-nitzschia*, alter their gene expression and metabolic processes in response to different Fe concentrations, and produce ferritin, a Fe-storage molecule that protects against oxidative stress (Marchetti et al., 2009; Armbrust, 2009). In particular, in *P. tricornutum*, uncomplexed Fe$^{3+}$ can be bound to a phytotransferrin (Iron Starvation-Induced Protein 2A, ISIP2A) at the plasma membrane surface in a CO$_3^{2-}$-dependent manner (Fig. 1.15), which confers the ability to concentrate Fe$^{3+}$ at the cell surface, as adaptation to the low-Fe environments of the contemporary ocean (Morrissey et al., 2015).

Some diatoms can use siderophore-chelated Fe$^{3+}$, such as *P. tricornutum*, in which hydroxamate siderophores are taken up without previous reduction by a very high-affinity mechanism, that involves a bond to the cell surface followed by endocytosis-mediated uptake and the delivery to the chloroplast (Kazamia et al., 2018). Moreover, the Iron Starvation-Induced Protein 1 (ISIP1) has been shown to be a diatom-specific protein, implying an evolutionary innovation in this group, and to putatively play an important role in the endocytosis of siderophores (Kazamia et al., 2018; Gao et al., 2021). Recently, Coale et al. (2019) identified a mechanism for which hydroxamate siderophore binds a Ferrochrome Binding Protein (FBP1) at the cell surface (Sutak et al., 2020) (Fig. 1.15).
T. pseudonana, which lacks FBP1 and ISIP1 genes, is unable to use siderophores as a source of Fe (Gao et al., 2021). As ISIP1-dependent siderophore uptake is nonreductive, as previously shown for ISIP2A, Fe$^{3+}$ taken up needs to be reduced: six *P. tricornutum* genes have been putatively annotated as Ferric Reductases (FREs), but only two (FRE1 and FRE2) encode two domains indicative of ferric reductase function (Gao et al., 2021). The *in silico* prediction suggested FRE2 is targeted to the chloroplast, while the localisation of FRE1 is still unassigned (Rastogi et al., 2018).

Other different pathways have been identified to model Fe uptake once it is bound by the cell membrane. Turnšek et al. (2021) identified two *P. tricornutum* proteins, PtTF.CATCH1 and PtF.CREG1, that co-localise with phytotransferrin ISIP2A, enabling the formulation of a model for Fe transport to the chloroplast: a vacuolar H$^+$-ATPase drives the endosome acidification and leads to Fe$^{3+}$ unbinding within the vesicle and its reduction to Fe$^{2+}$ inside the vesicle. Fe$^{2+}$ export from the chloroplast-docked vesicle likely occurs through a ZIP-type metal transporter and is sequestered by pTF.CATCH1 on the outer chloroplast envelope (Turnšek et al., 2021; Brownlee et al., 2022).

Among other micronutrients, low dissolved oceanic surface zinc (Zn) concentrations suggest that it may influence the growth of phytoplankton such as diatoms (Kellogg et al., 2022). A recent global proteomic analysis identified two proteins (ZCRP-A/B, Zn/Co Responsive Protein A/B) among four diatom species that became abundant under Zn/Co limitation, and whose characterisation suggests ZCRP-A could be a putative Zn/Co chaperone while ZCRP-B could be a membrane bound transport complex component (Kellogg et al., 2022).

### 1.3.4. Nitrogen transporters in diatoms

The macronutrient N is an important key molecule for all living organisms. It is supplied in several chemical forms within the marine environment. Dinitrogen (N$_2$) represents around 94% of total N, being the most abundant form in the atmosphere and dissolved in waters, that can be used by N$_2$-fixing bacteria but that is inaccessible to eukaryotic algae. However, symbiotic processes between heterocystous cyanobacteria and some species of diatoms were described, allowing diatoms to alternatively access to this unavailable dissolved gas (Caputo et al., 2019; Sanz-Luque et al., 2015). Usually, N can be incorporated in eukaryotic organisms from either organic or inorganic forms, whose availability is generally small and variable depending on environment (Sanz-Luque et al., 2015).
The major forms of inorganic N are NO$_3^-$, NO$_2^-$ and NH$_4^+$, while organic N sources are amino acids, proteins, nucleic acids, nucleotides, and urea.

NH$_4^+$ and NO$_3^-$ are the most commonly exploited N sources for phytoplankton, even if NH$_4^+$ is generally favoured. This is due to lower energetic costs for its acquisition: while NO$_3^-$ needs a reduction process to be assimilated, NH$_4^+$ can be more easily transported and used by the cell, in particular under N limiting conditions (Glibert et al., 2016). In particular, once NO$_3^-$ is transported into the cell, then a cytosolic Nitrate Reductase (NR) catalyses NO$_3^-$ reduction to NO$_2^-$, which subsequently is transported from the Nitrite Transporter 1 (NAR1) into the chloroplast, where the enzyme Nitrite Reductase (NiR) catalyses its reduction to NH$_4^+$. This step can take place in chloroplast or cytosol, and it is respectively catalysed by Fd-NiR and NAD(P)H-NiR (Armbrust et al., 2004) (Fig. 1.16). Finally, NH$_4^+$ is incorporated to carbon skeletons by transformation into glutamate, through the Glutamine Synthetase / Glutamine Oxoglutарат Amino Transferase or glutamate synthase (GS/GOGAT) cycle (Fig. 1.16): NH$_4^+$ is incorporated as the amide group of glutamine in a reaction involving glutamate and ATP (catalysed by GS) and then, the amide group is transferred reductively to a oxoglutarate to form two molecules of glutamate (Allen et al., 2011; Sanz-Luque et al., 2015). This NH$_4^+$ assimilation can be located in plastids (Fd-GOGAT / GSII) as well as in mitochondria (NAD(P)H-GOGAT / GSIII) (Allen et al., 2011) (Fig. 1.16).

In high NH$_4^+$ concentrations, NO$_3^-$ uptake is inhibited (Lomas and Glibert, 2000). This is a highly variable phenomenon because such inhibition can be activated by very high NH$_4^+$ concentrations (1μM) that are very uncommon in the open ocean, where concentration are widely lower, so these conditions are mainly confined to estuaries (Glibert et al., 2016). In particular, in the oceanic euphotic zone, closer to the surface and receiving enough light for photosynthesis, the estimated mean NO$_3^-$ concentration is about 7 μM while NH$_4^+$ and NO$_2^-$ ones are about 0.3 and 0.1 μM, respectively. Whereas in the aphotic zone, these values are 31, 0.01 and 0.006 μM for NO$_3^-$, NH$_4^+$ and NO$_2^-$, respectively (Sanz-Luque et al., 2015).

However, if cyanobacteria and many chlorophytes and dinoflagellates are better adapted to exploit NH$_4^+$, its preference has been recently questioned for diatoms, suggesting there can be diatoms that are NO$_3^-$ specialists (Glibert et al., 2016; Busseni et al., 2019). In fact, i) diatoms dominate in NO$_3^-$ enriched pelagic environments; ii) they use NO$_3^-$ even at very high concentration of NH$_4^+$ with stronger uptake rates and specific growth rates than
flagellates at a comparable substrate concentration; and iii) there are a large phylogenetic
distance and a difference in copies number between NO\textsubscript{3}\textsuperscript{-}
transporters of diatoms and other algae (Lomas and Glibert, 2000; Sanz-Luque et al., 2015; Armbrust et al., 2004;
Busseni, 2018; Rogato et al., 2015).

Diatoms are also able to exploit organic N forms, such as amino acids and urea, which contribute to around 30% of the global uptake (Allen et al., 2011), as they have a complete
ornithine-urea cycle, absent in plants and green algae but similar to the one present in animals. This metabolic pathway, inherited by the heterotrophic host of the secondary
endosymbiosis, has been proven to guarantee to diatoms a fast recovery after prolonged
N starvation and plays a role in nutrients transport between mitochondria, chloroplast and
cytoplasm (Allen et al., 2011; Smith et al., 2019; Busseni, 2018).

For all the different molecular forms of N sources, diatoms have specific transporters
(Busseni et al., 2019; Rogato et al., 2015): amino acid transporters, urea transporters (UT),
NH\textsubscript{4}\textsuperscript{+} transporters (AMT) and high/low affinity NO\textsubscript{3}\textsuperscript{-} transporters (NRT2/NPF) (Figs.
1.14, 1.15 and 1.16).

In culture, diatoms can use amino acids as a N source, but uptake and transport systems
for these are uncharacterised. The full suite of predicted diatom amino acid transporters
is very diverse, but most belong to the amino acid/auxin permease (AAAP) family (24
average per genome) (Smith and Allen, 2022).

Diatoms own two types of urea transporters: the first one, most similar to plant-like UTs,
is localised to the plasma membrane, while the other one, more similar to metazoan
kidney-type UTs and probably originated in the heterotrophic host (Fig. 1.14), has been
experimentally localised to mitochondria in \textit{P. tricornutum} (Smith et al., 2019).

As regard NH\textsubscript{4}\textsuperscript{+} uptake, plant AMT transporters can be divided in two subfamilies, AMT1
and AMT2. \textit{Arabidopsis thaliana}, for example, owns five AMT1 and one AMT2, while
eukaryotic phytoplanktonic organisms, on the basis of literature data, seem to have
multiple high affinity NH\textsubscript{4}\textsuperscript{+} transporters (AMT1). It has been reported that diatom
genomes encode for at least twice as many NH\textsubscript{4}\textsuperscript{+} transporter genes as the urea and NO\textsubscript{3}\textsuperscript{-}
transporters (Allen et al., 2006; Rogato et al., 2015). AMT1s are channel-like proteins
that act as NH\textsubscript{4}\textsuperscript{+} uniporters, NH\textsubscript{4}\textsuperscript{+}/H\textsuperscript{+} symporters or NH\textsubscript{3}/H\textsuperscript{+} co-transporters (Fig. 1.14).
While diatom AMT1 expression shows no clear modulation patterns related to
environmental conditions, their differential expression in response to a tight
compartmentalisation with NH$_4^+$ producing prokaryotes may suggest a cross-talk between bacteria-based NH$_4^+$-producing and diatom-based NH$_4^+$-assimilation pathways, similarly to that observed between land plants and root-associated microbiota (Busseni et al., 2019).

Then, NO$_3^-$ is taken into a cell via the NO$_3^-$ transporter proteins, which are divided into two categories: High-Affinity Transport System (HATS) operating at low NO$_3^-$ concentrations in the range of 0.2 mM, and Low-Affinity Transport System (LATS) active at higher NO$_3^-$ concentrations (more than 1 mM) (Rogato et al., 2015). Both perform proton (H$^+$) -coupled active transport in a symport mechanism that is driven by the pH gradients across membranes (Figs. 1.15 and 1.16), although some evidence suggests the use of Na$^+$ rather than H$^+$ in this symport mechanism for marine diatoms (Boyd and Gradmann, 1999).

The HATS is represented by the high-affinity NO$_3^-$ Transporters 2 (NRT2) gene family, which belongs to the Major Facilitator Superfamily (MPF) of transporters, and for which no substrate other than NO$_3^-$ has been identified so far. NRT2 association with the NO$_3^-$ Assimilation-Related protein (NAR2) is shown to be mandatory for NO$_3^-$ transport activity in higher plants and green algae (von Wittgenstein et al., 2014). However, no NAR2 proteins were found in public available diatoms genomes, indicating that their functional “high-affinity” NO$_3^-$ transport does not require the accessory protein NAR2 in diatoms (Rogato et al., 2015; Sanz-Luque et al., 2015). Considering that diatoms are unicellular organisms, the number of NRT2 members (for example six in *P. tricornutum* or five in *P. multistriata*) is high compared to plants (for example seven in *A. thaliana*), where NRT2 members are also involved in NO$_3^-$ distribution throughout different plant tissues. This can be justified by the importance of NO$_3^-$ as a major N source and the need to provide a quick response, in terms of nutrient uptake, to the constant and rapid nutritional changes that take place in the ocean environment (Rogato et al., 2015). Among six *P. tricornutum* NRT2, three have been predicted to localise to the plasma membrane, while the other three have been predicted to the vacuole membrane, called tonoplast (Busseni et al., 2019).

On the other hand, the LATS is represented by the low-affinity NO$_3^-$ Transporter 1/Peptide Transporter Family (NPF). In higher plants, NPFs includes more than 50 members (53 in *A. thaliana*) sub-classified in 8 phylogenetic clades (Léran et al., 2014). Differently from NRT2s, which are known to transport only NO$_3^-$, NPFs can transport
different molecules: from $\text{H}^+$ and short-chained di-/tri-peptides in bacterial Peptide Transporters (PTRs), also called Proton-coupled Oligopeptide Transporters (POTs), to a wide range of substrates in higher plants NPFs, such as amino acids, dicarboxylates, glucosinolates and phytohormones, as auxin and abscisic acid, beside $\text{NO}_3^-$ (Jørgensen et al., 2017; Kanno et al., 2012; Léran et al., 2020; Nour-Eldin et al., 2012).

The best studied one is NPF6.3 from *Arabidopsis* (also called AtNRT1.1) and represents a paradigmatic protein as it is considered a transceptor, with both sensor and carrier functions (Tsay et al., 2007; Gojon et al., 2011; Sanz-Luque et al., 2015). Moreover, AtNPF6.3 was actually shown to display an unusual dual-affinity transport activity, depending on phosphorylation of the Thr101 residue catalysed by the CIPK23 kinase: the phosphorylated form of AtNPF6.3 is a high-affinity carrier, whereas the non-phosphorylated form is a low-affinity one (Gojon et al., 2011; Ho et al., 2009). Probably because of the low $\text{NO}_3^-$ concentration in the surface ocean there are still no studies that characterised the low-affinity NPF family in marine diatoms and, in contrast with what observed in plants, Rogato et al. (2015) found only a small number of sequences encoding for putative NPF proteins in the genome of sequenced diatoms sharing a level of amino acid identity.
Figure 1.16. N and the diatom cell. Uptake from the extracellular environment occurs via different transporters. N internalised is then reduced through N assimilation pathways. NO$_3^-$: nitrate, NO$_2^-$: nitrite, NH$_4^+$: ammonium, Glu: glutamate, Gln: Glutamine. NRT/NPF: Nitrate Transporters, NR: Nitrate Reductase, NAR1: Nitrite Transporter, NiR: Nitrite Reductase, GS: Glutamine Synthetase, GOGAT: Glutamine Oxoglutarate Amino Transferase.

1.3.5. Ion balance in diatoms

Open ocean environments are typically defined by high Na$^+$ concentrations, around 450–500 mM, and a relatively stable, mildly alkaline pH, around 8.1. The high extracellular Na$^+$ concentration can be utilised by marine organisms to drive coupled transport processes across the plasma membrane (Taylor et al., 2012). In fact, many phytoplanktonic species are supposed to utilise a Na$^+$-based transport economy, with biophysical studies in marine diatoms indicating Na$^+$-coupled uptake of several nutrients, including NO$_3^-$, inorganic C and Si (Taylor et al., 2012; Hildebrand, 2005). This is a strong similarity with animal transport strategy, although there are still limited physiological or molecular evidence for a classical animal-like Na$^+$/K$^+$-type ATPase at the diatom plasma membrane.

H$^+$ have also a key role in coupling transport of molecules across membranes, and play many other roles in signalling and metabolic regulation. Many nutrient transporters have
been shown to co-transport anion together with H\(^{+}\), at the same time satisfying cell needs and balancing ion homeostasis (Léran et al., 2014; Taylor et al., 2012). Moreover, H\(^{+}\)-coupled nutrient transporters act together with plasma membrane voltage-gated H\(^{+}\) (H\(_{v}\)) channels, which have been proposed as a mechanism mediating rapid H\(^{+}\) efflux. H\(_{v}\) channels have been shown to have important functions in regulating intracellular pH, H\(^{+}\)-mediated signalling processes such as bioluminescence, nutrition and defence against grazers (Taylor et al., 2012). Many other metabolic processes can alter the intracellular H\(^{+}\) homeostasis: for example, HCO\(_{3}\)\(^{-}\) uptake for CCMs contributes to the cell net H\(^{+}\) consumption, while intracellular calcium carbonate precipitation from HCO\(_{3}\)\(^{-}\) in coccolithophores contributes to the net H\(^{+}\) generation (Taylor et al., 2012).

1.3.6. Nutrient sensing

To maintain homeostasis and respond to changing nutrient conditions, diatoms evolved mechanisms which allow them to sense extracellular nutrient concentrations and/or intracellular nutrient status and to enact the appropriate cellular responses through signal transduction (Smith and Allen, 2022). These responses aim to increasing uptake, remodelling photosynthesis and regulating metabolic enzymes concentration to drive shifting flux demands, such as controlling cellular recycling and growth.

Diatoms have evolved an array of nutrient transporters, including diverse families of channels, carriers, and pump proteins, that are modulated in accordance with nutrient demand and availability. In higher plants, some of these proteins, known as transceptors, have been demonstrated to play a signalling role in addition to their transport function (Gojon et al., 2011). In particular, plant NPFs can change their conformation and transport kinetics, according to NO\(_{3}\)\(^{-}\) availability in the environment (Fan et al., 2017; Jaubert et al., 2022). A similar function in nutrient sensing has been proposed for the SITs of *T. pseudonana* (Shrestha and Hildebrand, 2015).

Nutrient sensing acts through a cascading series of events, such as protein phosphorylation and Transcription Factors (TFs) activation, that allow cells to adjust their metabolism appropriately for the given stimulus (Smith and Allen, 2022). Nutrient sensing, signal transduction and regulatory pathways have been relatively well studied in other organisms, particularly for P deficiency and N metabolism in higher plants (Sanz-Luque et al., 2015), however little is still known for diatoms.
As regards to phosphorylation, a recent study integrating transcriptomic and phosphoproteomic analyses in *P. tricornutum* highlighted a big number of proteins conditionally phosphorylated as a function of light and N concentration, as well as a number of stress related kinases that might be responsible for phosphorylating them (Jaubert et al., 2022; Tan et al., 2020). Among them, calcium-dependent protein kinases, calcium calmodulin-dependent protein kinases and Ser/Thr kinases like TOR (Target Of Rapamycin), which play a central regulatory role in cell homeostasis, connecting qualitative and quantitative information, and which has been recently found in diatoms (Prioretti et al., 2017).

Use of molecular tools begun to elucidate some secondary messenger signals involved in nutrient sensing and signalling transduction in diatoms, such as the involvement of calcium (Ca$^{2+}$) as second messenger. Ca$^{2+}$ is well studied in animals (used for muscle contraction, fertilisation, and release of neurotransmitters) and plants (NO$_3^-$ sensing), but its existence in signal perception and transduction in diatoms is recent (Helliwell et al., 2021; Smith and Allen, 2022).

Furthermore, nitric oxide (NO) signalling has been shown to be involved in sensing infochemicals, and has been shown to be produced by diatoms mainly in response to grazing. In fact, NO is utilised in cell-to-cell communication, between stressed and bystander diatom cells, providing a stress surveillance system (Graff van Crevel, 2022).

Then, recent studies aimed to elucidate elements of signal transduction networks by identifying TFs that regulate nutrient-specific gene expression programs, mainly in *P. tricornutum*. A low P response regulatory TF, named PtPSR (Phosphorus Starvation Response), has been recently identified in *P. tricornutum* (Alipanah et al., 2015; Smith and Allen, 2022; Sharma et al., 2020). While, TFs that regulate the N-specific expression patterns observed in diatoms have been more elusive, because of the lack of clear homologs for N-specific regulatory TFs known in other organisms, with the only exception of *NirA* (Nitrate Assimilation TF), which may bind a motif identified in the promoters of genes upregulated specifically in response to NO$_3^-$ (Smith et al., 2019).

Diatoms live in a strongly variable environment and are continuously exposed to a multitude of different signals and stresses, so they are expected to implement interactions between regulatory pathways for the perception of stimuli and consequent responses (Jaubert et al., 2022). Much has still to be done to identify sensing and signalling...
components involved in nutrient responses in diatoms: for example, there is little knowledge about systems used by diatoms to sense nutrient status, to transmit these signals to the nucleus and to consequently regulate gene expression. Significant work needs to be done with the aim of identifying receptors and to characterise the signalling pathways downstream of these receptors. Moreover, subcellular localisation of many proteins is also unclear and often difficult to predict given the high degree of subcellular compartmentation observed in diatoms, potentially obfuscating detection of other novel adaptations diatoms possess (Jaubert et al., 2022; Smith and Allen, 2022).

1.4. Aim of the thesis

Low-affinity NO$_3^-$ transporters have been shown in other organisms to be active at relatively high NO$_3^-$ concentrations, in order of millimolar, that are usually not present in the ocean. So, it is really unexpected that diatoms evolved a transporter family working in conditions that are usually not present in their environment.

The general aim of my thesis was to assess the precise role of the low-affinity NO$_3^-$ transporters in diatoms (diNPFs) and to functionally characterise them in the model diatom *Phaeodactylum tricornutum*, integrating the use of eco-physiological and molecular approaches. This to understand the adaptation strategy of marine diatoms to sense and respond to variable and unfavourable environmental factors such as low nutrient condition. Moreover, this study could help to understand if and how NO$_3^-$ transporters can contribute to the regulation of the cell physiology in response to different signals.

The first questions that drove this PhD were: how many NPFs do diatoms have? How did diNPFs evolve compared to those of other organisms? What is their structure and where are they localised inside the cell?

To answer these questions, in Chapter 2, I investigated the evolution of these transporters in diatoms, in order to broaden the knowledge so far available for other organisms such as plants and bacteria, and to generate a complete picture of low-affinity NO$_3^-$ transporters evolution. Moreover, I studied the environmental conditions that regulate the expression levels of *diNPFs*, identifying interesting abiotic cues that could drive *diNPF* genes regulation. I predicted diNPFs molecular structures based on NPF crystal structures available for other organisms, comparing results with phylogenetic analyses and
identifying features and key residues important for the subsequent reconstruction of conformational functioning mechanism. Finally, I predicted the subcellular compartments where these transporters are localised, which allowed to propose new working hypotheses for diNPFs functional role.

After that, the proposed hypotheses on diNPFs functioning needed to be experimentally validated. So, *P. tricornutum* was chosen as model diatom for function characterisation of NPFs: its ecology is largely unknown, but it is well characterised in terms of functional genomics. Furthermore, innovative genomic and genetic resources became recently available, as CRISPR/Cas9 genome editing technology.

The *P. tricornutum* genome encodes for two NPFs, *PtNPF1* and *PtNPF2*.

The best way to study the function of a gene of interest is to switch on and switch off this gene, in order to detect phenotypic changes and identify altered metabolic processes in mutants compared to wild-type. So, in Chapter 3, I optimised a recent protocol to generate knock-out mutants through the CRISPR/Cas9 system in *P. tricornutum*, which allowed me to obtain single and double *PtNPFs* knock-out mutants. I screened them and I selected strains with biallelic mutations of *P. tricornutum* for further phenotypic characterisation. Then, I had the necessary tools to investigate the function of the two PtNPFs, as representatives of diNPFs.

In Chapter 4, I focused on PtNPF1. I generated transgenic lines expressing PtNPF1 in fusion with fluorescent tags which allowed to visualise its subcellular localisation, experimentally validating the prediction made in Chapter 2. This tool, together with organelle trackers available, also gave me the possibility to study morphological features of the vacuole in wild-type and mutated strains growing under different environmental conditions, such as N starvation. Then, phenotypic characterisation of overexpressing and knock-out strains was performed, through growth curves and biochemical analyses which helped to clarify the molecular mechanisms in which PtNPF1 is involved.

Similarly, in Chapter 5 I focused on PtNPF2. I generated overexpressing lines to observe PtNPF2 subcellular localisation and I exploited wild-type and mutated strains for further phenotypic characterisation. In particular, since previous results indicated a light response and fluorescent tags suggested a plastidial compartment involvement, I performed photosynthetic analyses on different strains to elucidate the possible role of PtNPF2.
The present work has the ambition of having a multidisciplinary nature, since it spans different fields including ecology, evolutionary biology, physiology, oceanography, molecular and computational biology. This study wants to provide a better understanding of the adaptation strategy of marine diatoms to deal with variable environmental factors such as nutrient availability.

The results will reveal a complex regulatory landscape, suggesting a functional diversification of NPF proteins that may contribute to the regulation of the cell physiology in response to diverse extracellular and intracellular signals.
Chapter 2: Diatom NPFs identification, evolution and structure


Phylogenetic analyses and correlation between TARA Oceans gene expression levels and environmental parameters were mainly performed by Dr. Luigi Caputi (Department of Integrative Marine Ecology at Stazione Zoologica Anton Dohrn of Naples, IT). Models of diNPF protein structures were built in collaboration with Dr. Antonella Longo (BioDiscovery Institute and Department of Biological Sciences, University of North Texas, Denton, USA).
**Abstract**

Nitrogen is an essential nutrient for all living organisms, and it is taken up from the environment in different organic or inorganic forms. Diatoms, the dominant component of phytoplankton in the ocean are continuously exposed to fluctuating environmental conditions and nutrient concentrations including nitrogen.

As other organisms, diatoms rely on a range of transmembrane transporters for nitrogen uptake, among which the low-affinity Nitrate Transporter 1/Peptide Transporter Family (NPF). These transporters, well characterised in other organisms such as bacteria and higher plants, have the particular feature of being able to recognise a large number of different substrates. However, little is still known about diatom NPFs.

In this Chapter, diatom NPFs were identified and investigated through an integrated approach which combined omics, phylogenetic, structural and expression analyses. Diatom NPF genes diverge to produce two distinct clades with strong sequence and structural homology with either bacterial or plant NPFs, with different predicted subcellular localisation, suggesting that their divergence could result in functional diversification and revealing an unexpected complexity of these transporters.
2.1. Introduction

In the ocean, the nutrient availability strongly fluctuates in both time and space. In order to cope with these continuous changes and to maximise their utilisation, living organisms need efficient sensing and uptake systems that rapidly check the external concentrations of single nutrients for their transport and assimilation, as well as intracellular mechanisms which allow their storage and reallocation (Gojon et al., 2011).

Among all nutrients, nitrogen (N) is one of the most important for photosynthetic organisms, that can be taken up from the environment exists in different inorganic or organic forms, as nitrate (NO$_3^-$), nitrite (NO$_2^-$), ammonium (NH$_4^+$) and urea.

In particular, NO$_3^-$ represents the major bioavailable N source for most photosynthetic organisms, being necessary as both an essential nutrient and a signal molecule involved in many metabolic pathways, growth, development, and adaptation to various N conditions. However, its availability is generally low and intermittent in the ocean (Vidal and Gutiérrez, 2008; Zhang et al., 2018).

NO$_3^-$ is taken up by cells via the NO$_3^-$ transporter proteins. The NO$_3^-$ transport system consists of the High-Affinity Transport System (HATS), which includes the Nitrate Transporters 2 (NRT2) family active at low NO$_3^-$ concentration, and the Low-Affinity Transport System (LATS), which includes Nitrate Transporter 1 (NRT1) gene family, recently renamed as the NRT1/PTR Family (NPF), active at high NO$_3^-$ concentration (Léran et al., 2014).

NPFs are phylogenetically related to the family of Peptide Transporters (PTRs) or Proton-coupled Oligopeptide Transporters (POTs) that are evolutionarily conserved in archaea, bacteria, fungi, yeast, algae and higher plants (MacGregor et al., 2013). NPFs, like POTs, PTRs, and NRT1s, belong to the Major Facilitator Superfamily (MFS) of secondary active transporters, that includes facilitators, symporters, and antiporters, and that move substrates across membranes via facilitated diffusion, co-transport, or exchange, respectively (Yan et al., 2013). A remarkable feature of POT/PTR/NPF family is the diversity and extent of natural substrates they recognise. These range from short chained di- and tripeptides in bacteria, fungi and mammals to a wide variety of molecules in higher plants, such as NO$_3^-$ (Parker et al., 2017), NO$_2^-$, di-/ tripeptides, amino acids, dicarboxylates, glucosinolates (Jørgensen et al., 2017; Nour-Eldin et al., 2012) and
phytohormones, as auxin (IAA), gibberellic acid (GA) (Chiba et al., 2015; Wulff et al., 2019) and abscisic acid (ABA) (Kanno et al., 2012; Léran et al., 2020).

However, individual members can show tight substrate selectivity; for example, the plant proton (H⁺)-coupled NO₃⁻ transporter AtNPF6.3/NRT1.1 can recognise NO₃⁻ and IAA, but not peptides (Parker and Newstead, 2014; Corratgé-Faillie and Lacombe, 2017). Some peptide transporters, however, can recognise more than 8000 different di- and tripeptides, which are themselves chemically diverse (Fowler et al., 2015; Parker et al., 2017).

Although NPFs have evolved to transport a variety of substrates, their overall structures are highly conserved. The crystal structures for one plant NPFs and several bacterial POTs were all solved in the inward open conformation, either as apo-proteins or bound to NO₃⁻ or peptides (Sun and Zheng, 2015; Boggavarapu et al., 2015; Parker and Newstead, 2014; Parker et al., 2017; Quistgaard et al., 2016; Sun et al., 2014). Structural analyses showed that NPFs and POTs are integral membrane proteins, folded into 12 transmembrane helices (TMH) organised into two bundles, the N-terminal domain that includes TMH1 - 6, and the C-terminal domains with TMH7 - 12. Both N- and C-termini are located at the cytosolic membrane side (Longo et al., 2018). The substrate-binding site is located in a clearly defined cavity that extends from the middle of the proteins towards the intracellular space.

Moreover, the transport mechanism is expected to be conserved independently from the substrate transported. It has been established that POTs and the majority of NPFs, which are substrate-H⁺ symporters, employ an alternating access mechanism. In this process, two H⁺ attach to charged amino acids in the TMH1 ExxER motif to initiate transport when the protein is in the outward open conformation. A salt bridge between the oppositely charged residues on TMH4 and TMH10, which also functions as an intracellular gate, stabilises this conformation (Newstead, 2015, 2017; Longo et al., 2018). A particular biochemical property of the POT/PTR/NPF family is their ability to recognise diverse ligands while retaining a strict requirement to couple transport to the electrochemical proton gradient (Longo et al., 2018).

The first transporter of NO₃⁻ identified in plants was AtNPF6.3/NRT1.1 (Tsay et al., 1993), localised on the plasma membrane. It was first identified as an inflow carrier involved in the LATS of NO₃⁻ in roots, but later it was reclassified as a dual-affinity NO₃⁻ transceptor (transporter/receptor), working in NO₃⁻ dependent regulation: the
phosphorylated form of the transporter demonstrated a high affinity activity, whereas the non-phosphorylated form demonstrated a low-affinity activity (Zhang et al., 2018; Liu and Tsay, 2003). For it and *Oriza sativa* OsNPF6.5, a dual-affinity transporter for NO$_3^-$ has been demonstrated, and in several cases the capacity of low-affinity NPFs to transport NO$_3^-$ even in a range of concentrations extending from 10 μM to 500 μM has been reported (Valkov et al., 2017; Bagchi et al., 2012; Hu et al., 2015; Wang et al., 2020).

In bacteria, the evolutionary related POTs, also called PTRs are responsible for the uptake and transport of small peptides (Yan et al., 2013; Newstead, 2017). POTs are believed to function by an alternate access mechanism involving gated transitions between inward open, occluded and outward open conformational states (Newstead, 2015), which is essentially similar to that of any other MFS transporter (Quistgaard et al., 2016).

In bacteria and animals, this kind of transporters are present in a low number of copies (1 to 4 for bacteria; 1 in yeast; 3 in *Drosophila melanogaster* and *Caenothabditis elegans; 4 in man); in algae, NPFs are present in 1 (*Chlamydomonas reinhardtii* and *Micromonas pusilla*) or 2 copies (*Coccomyxa subellipsoida* and *Chlorella variabilis*), with some algae that do not present these transporters at all, like *Chlorella paradoxa*. By contrast, in the plant genomes there is a high number of NPFs, from the lowest of 20 members in *Physcomitrella patens* up to 139 in *Malus domestica* (Léran et al., 2014; Sanz-Luque et al., 2015).

While in terrestrial multicellular phototrophs the localisation and regulation of NPF proteins have been widely studied, for marine unicellular phototrophs, whose environment is characterised by low and fluctuating concentrations of their substrates (Garcia *et al.*, 2010), exploration is still in its infancy.

This Chapter is focused on NPFs in diatoms (diNPFs), the dominant component of phytoplankton, contributing up to 40% of marine primary production and most of the CO$_2$ export (drawdown and sinking) from the atmosphere, of the order of 50% (Hildebrand et al., 2012). Characterised by complex evolutionary history, diatoms developed a unique metabolism and subcellular organisation, allowing them to greatly adapt to changes in environmental stress conditions (Armbrust, 2009). Diatoms are indeed considered as opportunistic (*r*-strategist) able to grow rapidly and store resources when they are in surplus and to re-utilise them during unfavourable periods, so outcompeting other phytoplankton in strongly fluctuating waters (Lagus et al., 2004; Margalef, 1978).
In the last years, N metabolism in diatoms has inspired many studies: in fact, besides being one of the most important macronutrients for the growth of diatoms, together with silicon and phosphorus (Chen et al., 2018), it constitutes the skeleton of the fundamental molecules that make up organisms, such as nucleic acids and amino acids (Stankiewicz and van Bergen, 1998). Furthermore, in diatoms, as well as in many other microalgae, the condition of N starvation is known to cause lipid accumulation, molecules interesting from a biotechnological point of view for biofuel production (Zulu et al., 2018). Analyses have demonstrated the complex chimeric nature of diatom genomes (Rastogi et al., 2020), and studies have revealed that their overall network of N metabolism is a combination of plant-like, animal-like, and bacterial traits (Armbrust et al., 2004; Bowler et al., 2008; Allen et al., 2011). Mainly for these reason, different studies on N metabolism have started to guide the understanding of how diatom cells utilise N for growth (Kang and Rynearson, 2019; Yang et al., 2014), N metabolism inside the cells (Allen et al., 2011; McCarthy et al., 2017) and N fluxes between different organelles (Smith et al., 2019), trying to clarify N metabolic network connectivity in diatoms.

However, little is still known about N transport, including NO$_3^-$ sensing, uptake and molecular mechanism of transport across diatom membranes. In particular, the role of diNPFs remains unclear, and, unlike plants, where NPF have been well studied and classified, for diNPFs there are no functional characterisation works.

Moreover, the presence of low-affinity NO$_3^-$ transporters in marine species is actually puzzling, since in the oceans NO$_3^-$ concentrations are always lower than the approximate concentration (1 mM) at which NPFs have shown to become efficient in other organisms. In particular, NO$_3^-$ oceanic concentrations are of the order of micromolar, while a big number of higher plants NPFs become active at millimolar concentrations. This suggests that NPFs in marine organisms are either used for internal transport where local NO$_3^-$ levels are higher, or needed to transport different substrates than NO$_3^-$ (Glibert, 2016; Busseni et al., 2019).

In this context, diNPFs evolutionary history, their structure and their role need to be clarified. To address these issues, data coming from fully sequenced diatom genomes (Rogato et al., 2015) were integrated with: the TARA Oceans dataset (de Vargas et al., 2015; Carradec et al., 2018), similarly to what previously done for NRT2 and AMT in Busseni et al. (2019), which provides a unique combination of environmental, metagenome and metatranscriptome data, and the Marine Microbial Eukaryotic
Transcriptome Sequencing Project (MMETSP) dataset, which provides over 650 assembled, annotated and public transcriptomes (Keeling et al., 2014).

These results contributed to draw the first overview of NPFs genes in marine diatoms, revealing an unexpected evolutionary complexity of NPF transporters and laying the foundation for further function analyses on this puzzling transporter family.

2.2. Materials and Methods

2.2.1. diNPFs identification in public genomes, TARA Oceans and MMETPS databases

A first search of the diNPFs sequences was performed using sequences described by Rogato et al. (2015), which included a list of 13 diNPFs, from four diatom species, whose genomes were publicly available at the time (Rogato et al., 2015).

The translated protein sequences were used as queries in BlastP searches with default parameters against the following databases: JGI Genome Portal (https://genome.jgi.doe.gov/portal/), Ensembl Protist (https://protists.ensembl.org/index.html) and PLAZA Diatoms 1.0 (https://bioinformatics.psb.ugent.be/plaza/versions/plaza_diatoms_01/). Subsequently, a BlastP was performed against the peptides deriving from the genome of Arabidopsis thaliana, using TAIR (https://www.arabidopsis.org/).

Then an extensive search for putative diNPF genes was performed in both the Marine Atlas of TARA Oceans Unigenes (MATOU) database (http://tara-oceans.mio.osupytheas.fr/ocean-gene-atlas/) (Genoscope Technical Team et al., 2017; Carradec et al., 2018) and the MMETSP database (https://www.imicrobe.us/#/search/mmetsp) (Keeling et al., 2014), using diNPF sequences from the previous report (Rogato et al., 2015) and new diNPF homologues found in public genomes (Armbrust et al., 2004; Bowler et al., 2008; Lommer et al., 2012; Tanaka et al., 2015; Traller et al., 2016; Basu et al., 2017; Mock et al., 2017; Osuna-Cruz et al., 2020) as queries. BlastP against the TARA Ocean gene catalogue was performed using an in-house developed Blast platform, setting an Expected Threshold of 1 ∙ 10^{-5}.

The TARA Ocean gene catalogue was downloaded at http://www.genoscope.cns.fr/tara/.
Similarly, BlastP against the MMETSP peptides database was performed using an in-house developed platform.

2.2.2. diNPFs alignment

Sequences alignment was performed by subsequent rounds using the MAFFT (v.7) software at the MAFFT web portal. Initially, the experimental algorithm for a large number of short and similar sequences (https://mafft.cbrc.jp/alignment/server/large.html?aug31) was used. At each alignment round, sequences not satisfying some conditions were deleted from the alignment: in detail, sequences with length minor than 75 amino acids, which could be incomplete, and sequences with low quality scores were not more considered during the next steps. Moreover, sequences similar more than 75% with the reference genes, were deleted from the alignment. This arbitrary cut-off was chosen to avoid overlapping between retrieved sequences from dataset and queries, and represents a compromise between a representation of biodiversity and the complexity of data management. Subsequently, sequences alignment was performed at the MAFFT (v.7) web portal using standard options (https://mafft.cbrc.jp/alignment/server/). TrimAl (http://trimal.cgenomics.org/) was used to remove spurious sequences or poorly aligned regions from the alignment. Finally, the final alignment was obtained in-house using the ClustalW algorithm as implemented in Bioedit (v7.0.5.3). Alignment statistics were calculated using MEGA (v10.0.5).

2.2.3. Phylogenetic analyses

Different NPF sequences from bacteria, fungi, protists, plants and animal species were retrieved from public databases and added to the initial alignment, in order to construct a robust multi-kingdom NPFs alignment for phylogenetic purposes (the complete set of sequences is presented in Supplementary Table S2.1). ProtTest as implemented in MEGA (v10.0.5) was used in order to find the Best Protein Fitting Model (BPFM) for the given alignment: LG + γ + I + F, where γ=1.27, was the selected BPFM. Maximum Likelihood phylogeny was inferred using IQ-TREE (v1.6.12) for Windows system, with 1000 BS replicates. Baesyan Inference was performed using MrBayes (v3.2.7), using two chains. Baesyan Inference required 28 000 000 generations to reach an average standard deviation of split frequencies <0.01.
2.2.4. Global meta-omics of MATOU diNPFs

All analyses in this section have been performed in R (v3.4.4-win). Pearson correlation and relative $p$-values were calculated in the R package ggplot2.

2.2.5. Culture conditions

An axenic CCMP632 strain of *Phaeodactylum tricornutum* Bohlin was obtained from the Provasoli-Guillard National Centre for Culture of Marine Phytoplankton. It corresponds to the ecotype Pt1 (De Martino et al., 2007; Rastogi et al., 2020). The culture was maintained in autoclaved 0.22 μm filtered F/2 medium without silica (Guillard, 1975) at 18 °C under white fluorescent lights (90 μmol m$^{-2}$ s$^{-1}$), with a 12:12 hours dark-light photoperiod.

Cells were grown in different conditions for the detection of *PtNPFs* mRNA abundance and collected during the exponential growth phase. As main control, *P. tricornutum* Pt1 wild-type was grown in F/2 medium without silicate (Guillard, 1975), so with a NaNO$_3$ concentration of 882 μM as N source and a pH of 8.0. The experiments shown in Figures 2.4 were achieved by diluting *P. tricornutum* cells in F/2 medium containing 882μM NaNO$_3$ in their respective media with 50 μM NaNO$_3$, 882μM NH$_4$NO$_3$ or 882μM NH$_4$Cl; harvested after 5 days and 2 hours after the onset of light, and then collected for RNA extraction. Then, *PtNPFs* mRNA abundance was tested on the *P. tricornutum* Pt4 ecotype, cells grown in F/2 medium containing 882μM NaNO$_3$. Cells were collected one hour before and two hours after the onset of light (Fig. 2.4). Finally, cells were grown respectively at pH 7.0, 8.0 and 9.0 and harvested after 5 days to perform the analyses (Fig. 2.4).

2.2.6. RNA extraction, primer design and qPCR

RNA extraction and quantitative Real-Time reverse-transcription Polymerase Chain Reaction (qRT-PCR or qPCR) were performed as described in Russo et al. (2015). In detail, culture volumes corresponding to around 10$^8$ cells were pelleted through centrifugation for 15 minutes at 4000 g and 4°C. After being washed with PBS (1X) to completely eliminate sea salts, pellets were stored at -80°C until RNA extraction. So, total RNA was isolated using 1.5 ml TRI Reagent (SIGMA Life Science), and after 5 minutes at room temperature, 300 μl chloroform were added for homogenisation. Sample tubes were shaken by hand for 15 seconds and incubated 15 minutes at room temperature. After a centrifuge at 4°C, 12 000 g for 15 minutes, the upper colourless phase was recovered.
and mixed with 1 volume of cold isopropanol. Tubes were inverted several times and incubated 30 minutes on ice. After a centrifuge at 4°C, 12 000 g for 15 minutes, supernatant was discarded and 1 ml ethanol 75% was added. After a centrifuge at 4°C, 7600 g for 15 minutes, ethanol was discarded, RNA pellets were dried and then resuspended in DEPC water.

RNA concentration was determined using a Nanodrop™ (ND 1000 Spectrophotometer) and qualitatively estimated by gel electrophoresis (1% agarose w/v). Two hundred nanograms of total extracted RNA were reverse transcribed with QuantiTect Reverse Transcription Kit (Qiagen) according to the manufacturer’s instructions.

To assess cDNA quality and absence of gDNA contamination, Reverse Transcriptase (RT) -PCRs were run with control PtNOA gene (Nitric Oxide Associated protein, Phat3_J40200) using intron spanning primers PtNOA_for and PtNOA_rev (Table 2.1).

The qPCR reaction was performed according to Siaut et al. (2007). 1 μl of a 1:2 dilution of cDNA was used as template to amplify the PtNPF transcripts using 0.4 μM final concentration of the primers. The PCR primer sets were specifically designed to amplify a 150-250-bp fragment from the two PtNPF genes (Table 2.1). Primers were designed using the Primer3 software ([http://primer3.ut.ee/](http://primer3.ut.ee/)). RPS (Ribosomal Protein Small subunit 30S; ID 10847) was used as reference gene for qPCR data normalisation (Table 2.1) (Siaut et al., 2007).

<table>
<thead>
<tr>
<th>Target gene</th>
<th>Gene ID</th>
<th>Primer name</th>
<th>Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PtNOA</td>
<td>40200</td>
<td>PtNOA_exp_for</td>
<td>5'-CAGTTACTGACCCCCGAAGA-3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PtNOA_exp_rev</td>
<td>5'-AACGCACCTTCCGGTAGAGA-3'</td>
</tr>
<tr>
<td>PtNPFs</td>
<td>47148</td>
<td>PtNPFl_exp_for</td>
<td>5'-TTACGTGATGGCTCGTCCA-3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PtNPFl_exp_rev</td>
<td>5'-GGTCCGGCTTATATAACAGA-3'</td>
</tr>
<tr>
<td></td>
<td>47218</td>
<td>PtNPF2_exp_for</td>
<td>5'-CTACGAAGTCGCTTACCG-3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PtNPF2_exp_rev</td>
<td>5'-ATCTCAACCAGCGTGATAC-3'</td>
</tr>
<tr>
<td>PtRPS</td>
<td>10847</td>
<td>RPS_for</td>
<td>5'-GTGCAAGAGACCGGACATACC-3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RPS_rev</td>
<td>5'-CGAAGTCAACCAGGAAAACCA-3'</td>
</tr>
</tbody>
</table>
qPCR amplification was performed using Light Cycler™ 480 SYBR Green I Master 2X (Roche) in a final volume of 10 μl. Each reaction was performed in triplicate for each gene in each sample using 384-well plates (BioRad) in the ViiA™ 7 Real-Time PCR System (Applied Biosystems). Thermocycler settings were 95°C for 10 min, 40 cycles at 95°C for 1 sec and 60°C for 20 sec, a gradient from 60°C to 95°C for 15 min, to finish with 5 min at 72°C.

Data obtained were processed with the ViiA™ 7 Real-Time PCR system software. Fold-changes were obtained with the Relative Expression Software Tool-Multiple Condition Solver (REST-MCS) (Pfaffl et al., 2002) and values above ±2 were considered significative.

2.2.7. In silico analysis of the diNPFs non-coding sequences.

The 5′-flanking regions (500-1000 bp upstream sequence between the coding gene of interest and the upstream gene) of diNPF genes were obtained for diatom species whose genomes were available. MEME Suite (v5.1.1) (http://meme-suite.org/tools/meme) was used to screen for motifs with known transcription binding sites (Bailey et al., 2009, 2015), with a p-value cut-off of 10^{-3}. The top motifs were submitted to TOMTOM (v5.1.1) (http://meme-suite.org/tools/tomtom) with default parameters and searched against the Eukaryotic DNA, JASPAR CORE and UniPROBE Mouse databases.

FIMO (v5.1.1) (http://meme-suite.org/tools/fimo) was used to research specifically in the upstream sequences of diNPF genes, two motifs previously identified by Smith et al. (2019), which are responsible to bind helix-turn-helix-type transcription factors and regulate the expression of numerous High-Nitrate-Sensitive genes.

2.2.8. Structural prediction of diNPFs

Multiple sequence alignments of amino acid sequences were performed using the Clustal Omega program via the Web Services interface at the European Bioinformatics Institute (EMBL-EBI) (Sievers et al., 2011; McWilliam et al., 2013).

Structural data are available for one plant NPF, Arabidopsis thaliana NPF6.3 (Parker and Newstead, 2014; Sun et al., 2014) and several of the evolutionary related bacterial POTs (Newstead, 2011; Solcan et al., 2012; Doki et al., 2013; Guettou et al., 2014; Quistgaard et al., 2017; Lyons et al., 2014; Zhao et al., 2014; Ho and Frommer, 2014; Parker et al., 2017). Structural models of diatom NPFs were built by homology using the coordinates
from the crystal structures of plant NPF or bacterial POTs as templates. Homology models were obtained using the Swiss-Model Workspace (Arnold et al., 2006; Biasini et al., 2014). PyMol was used for molecular visualization (The PyMOL Molecular Graphics System, Schrödinger, LLC).

2.2.9. Prediction of subcellular localisation

The diNPFs subcellular localisation was predicted by running all the diNPF sequences, obtained from public diatom genomes and from TARA Oceans and MMETSP databases, through the LocTree3 software, PSI-BLAST (https://rostlab.org/services/loctree3/), pipeline PredictProtein (Yachdav et al., 2014).

Obtained results were compared with other software, such as WolfPSort (https://wolfsort.hgc.jp/), which uses an algorithm based on sequence homology (data not shown) and which confirmed the results. The subcellular localisation for the full-length sequences was also predicted through the SignalP software (http://www.cbs.dtu.dk/services/SignalP/), that uses an algorithm based on signal peptide research.

2.3. Results

2.3.1. diNPF genes in diatom genome, TARA Oceans and MMETPS databases

The NPF family of low affinity NO₃⁻ transporters in plants includes from as low as 20 members in Physcomitrella patens to 53 members in A. thaliana and more than 100 in Glycine max (Léran et al., 2014; Longo et al., 2018). Phylogenetic analysis allowed the identification of eight clades, each defining a distinct subfamily (Léran et al., 2014).

Using eight different A. thaliana NPF protein sequences, one for each identified clade (Léran et al., 2014), and already known diNPF sequences identified in a previous work (Rogato et al., 2015), diNPFs were searched in all the diatom genomes sequenced (Lommer et al., 2012; Basu et al., 2017; Mock et al., 2017; Osuna-Cruz et al., 2020; Tanaka et al., 2015; Traller et al., 2016; Armbrust et al., 2004; Bowler et al., 2008), through a BlastP searches in different public databases. Homologues were found in all diatom species analysed.
First of all, it was possible to make some changes to the already published data in Rogato et al. (2015), thanks to recent genomes annotation update, as can be seen in detail in the Table 2.2. In particular, the chromosomal location of *P. tricornutum* 47218 and *P. multiseries* 226109 was changed. *T. pseudonana* 269333 was updated not only in its chromosomal location but also in its sequence ID, which was previously 7452. One of the retrieved *F. cylindrus* NPF sequences (ID 256377) was not the complete sequence and was not analysed more in detail, as previously reported in Rogato et al. (2015). Moreover, new public genomes allowed to identify NPFs in other diatom species which, similarly to other diNPFs, reported common features, such as the presence of 12 Transmembrane Helices (TMHs) in their structure, and a high homology level with plant NPFs and a small number of introns, between 0 and 2, with the only exception of *P. multiseries* 190665 which owns 5 introns.

As to diatoms, most of the analysed species, such as *P. tricornutum, P. multiseries, P. multistriata, T. pseudonana, T. oceanica, S. marinoi* and *C. cryptica*, own two NPF genes, while *F. cylindrus, F. solaris* and *S. robusta* have more than two NPFs, probably due to duplication phenomena.

In particular, in the cold-adapted diatom *F. cylindrus*, two couples of the retrieved sequences (IDs 136520/256377 and 186175/204239) are annotated as allelic variants (Table 2.2) and have respectively 97% and 98% of sequence identity (Rogato et al., 2015; Mock et al., 2017). Approximately 25% of the diploid *F. cylindrus* genome consists of genetic loci with alleles, that are highly divergent (15 Mb of the total genome size of 61 Mb) and probably involved in adaptation to environmental fluctuations in the Southern Ocean (Mock et al., 2017).

Also *F. solaris* owns three couples of sequences (IDs 15278/26459, 17535/17972 and 20187/25493) annotated as allelic variants, with a sequence identity of respectively 87%, 98% and 96%. It is an allodiploid diatom (Tanaka et al., 2015), with high number of Transposable Elements (TEs) indicating a genome reorganisation and altered gene expression proper to allopolyploid species (Parisod et al., 2010; Tanaka et al., 2015).

Each of the diNPF sequences shows a high level of homology, between 22% and 42%, with different *A. thaliana* NPFs. The most represented are AtNPFs belonging to Clade 8 (Léran et al., 2014), with an average value of 27% amino acid identity (Table 2.2, only the top hit in the *Arabidopsis* database is shown). In *Arabidopsis*, this NPF Clade includes mainly dipeptide transporters localised on plasma membrane or on vacuole membrane;
however in rice they are low-affinity NO$_3^-$ transporter (Léran et al., 2014). Other plant NPFs highly represented in this analysis belong to Clades 5 and 6, which transport mainly NO$_3^-$, but which have been shown to transport also bigger molecules as phytohormones (Léran et al., 2014).

Data represented by the chromosomal location and the number of introns of the NPF sequences of *Skeletonema marinoi* were missing. In fact, the amino acid sequences related to this species NPFs were identified using only transcriptomic data (Amato et al., 2018), because the relative genome is not yet assembled and published.

The same research was performed thanks to the transcriptome of *Chaetoceros decipiens* (Amato et al., 2017): *Chaetoceros* is one of the largest genera of diatoms in the marine phytoplankton, and its many species are widely distributed, some even cosmopolitan (Li et al., 2017b). While Busseni et al. (2019) previously identified NRT2 and AMT transporters for this genus, here interestingly, no NPFs was detected for *C. decipiens*.

**Table 2.2.** List and properties of diNPFs, modified and integrated from Rogato et al. (2015). Superscript letters in the protein ID column denote allelic pairs. The *F. cylindrus* 256377 sequence was incomplete and was not analysed more in detail. TMs: transmembrane domains.
<table>
<thead>
<tr>
<th>Diatom</th>
<th>Protein ID</th>
<th>Chromosomal location</th>
<th>n° introns</th>
<th>aa length</th>
<th>n° TMs</th>
<th>Top BlastP hit in <em>A. thaliana</em></th>
<th>% identity</th>
<th>Genome reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>P. tricornutum</em></td>
<td>47148</td>
<td>chr_12: 617225-619552 (−)</td>
<td>0</td>
<td>775</td>
<td>12</td>
<td>AtNPF5.2 AT5G46050</td>
<td>33</td>
<td>(Bowler et al., 2008)</td>
</tr>
<tr>
<td></td>
<td>47218</td>
<td>chr_12: 843193-845146 (+)</td>
<td>0</td>
<td>650</td>
<td>12</td>
<td>AtNPF8.2 AT5G01180</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td><em>T. pseudonana</em></td>
<td>4104</td>
<td>chr_4: 182289-184968 (+)</td>
<td>3</td>
<td>765</td>
<td>12</td>
<td>AtNPF5.14 AT1G72120</td>
<td>38</td>
<td>(Armbrust et al., 2004)</td>
</tr>
<tr>
<td></td>
<td>269333</td>
<td>chr8: 269367-271586 (+)</td>
<td>3</td>
<td>592</td>
<td>12</td>
<td>AtNPF5.7 AT3G53960</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td><em>T. oceanica</em></td>
<td>32021</td>
<td>scaffold_5551: 489-2290 (−)</td>
<td>2</td>
<td>547</td>
<td>12</td>
<td>AtNPF8.2 AT5G01180</td>
<td>36</td>
<td>(Lommer et al., 2012)</td>
</tr>
<tr>
<td></td>
<td>14591</td>
<td>scaffold_23330: 1-700 (−)</td>
<td>0</td>
<td>233</td>
<td>5</td>
<td>AtNPF5.10 AT1G22540</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td><em>P. multiseries</em></td>
<td>190665</td>
<td>scaffold_103: 217004-220392 (+)</td>
<td>5</td>
<td>688</td>
<td>12</td>
<td>AtNPF2.10 AT3G47960</td>
<td>54</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>226109</td>
<td>scaffold_1727: 1943-4182 (−)</td>
<td>2</td>
<td>619</td>
<td>11</td>
<td>AtNPF6.4 AT3G21670</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td><em>P. multistriata</em></td>
<td>7930</td>
<td>chr18: 108896-111744 (+)</td>
<td>0</td>
<td>791</td>
<td>12</td>
<td>AtNPF4.7 AT5G62730</td>
<td>30</td>
<td>(Basu et al., 2017)</td>
</tr>
<tr>
<td></td>
<td>12290</td>
<td>chr_31: 54727-57165 (−)</td>
<td>2</td>
<td>654</td>
<td>12</td>
<td>AtNPF8.3 AT2G02040</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td><em>F. cylindrus</em></td>
<td>186175</td>
<td>scaffold_6: 1219622-1221532 (+)</td>
<td>2</td>
<td>565</td>
<td>12</td>
<td>AtNPF5.11 AT1G72130</td>
<td>39</td>
<td>(Mock et al., 2017)</td>
</tr>
<tr>
<td></td>
<td>204239</td>
<td>scaffold_110: 45530-47443 (+)</td>
<td>2</td>
<td>566</td>
<td>12</td>
<td>AtNPF5.11 AT1G72130</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td></td>
<td>136520</td>
<td>scaffold_9: 1493580-1495600 (−)</td>
<td>2</td>
<td>571</td>
<td>12</td>
<td>AtNPF6.4 AT3G21670</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>256377</td>
<td>scaffold_75: 193898-195223 (−)</td>
<td>2</td>
<td>441</td>
<td>6</td>
<td>AtNPF2.5 AT3G45710</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td></td>
<td>171976</td>
<td>scaffold_11: 1319758-1322441 (−)</td>
<td>2</td>
<td>589</td>
<td>12</td>
<td>AtNPF8.1 AT3G54140</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>147192</td>
<td>scaffold_15: 1337290-1339086 (+)</td>
<td>2</td>
<td>527</td>
<td>12</td>
<td>AtNPF6.4 AT3G21670</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200740</td>
<td>scaffold_56: 206598-208532 (+)</td>
<td>2</td>
<td>585</td>
<td>12</td>
<td>AtNPF8.1 AT3G54140</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>Diatom</td>
<td>Protein ID</td>
<td>Chromosomal location</td>
<td>n° introns</td>
<td>aa length</td>
<td>n° TM</td>
<td>Top BlastP hit in <em>A. thaliana</em></td>
<td>% identity</td>
<td>Genome reference</td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>----------------------</td>
<td>------------</td>
<td>-----------</td>
<td>-------</td>
<td>--------------------------------</td>
<td>------------</td>
<td>-----------------</td>
</tr>
<tr>
<td><em>F. solaris</em></td>
<td>15278&lt;sup&gt;c&lt;/sup&gt;</td>
<td>scaffold_89: 119806-123050 (+)</td>
<td>1</td>
<td>552</td>
<td>12</td>
<td>AtNPF8.3 AT2G02040</td>
<td>31</td>
<td>(Tanaka et al., 2015)</td>
</tr>
<tr>
<td></td>
<td>26459&lt;sup&gt;c&lt;/sup&gt;</td>
<td>scaffold_246: 21487-25295 (−)</td>
<td>1</td>
<td>561</td>
<td>12</td>
<td>AtNPF5.4 AT3G54450</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17535&lt;sup&gt;d&lt;/sup&gt;</td>
<td>scaffold_118: 68724-71015 (−)</td>
<td>0</td>
<td>614</td>
<td>12</td>
<td>AtNPF8.2 AT5G01180</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17972&lt;sup&gt;d&lt;/sup&gt;</td>
<td>scaffold_123: 425437-427851 (−)</td>
<td>0</td>
<td>684</td>
<td>12</td>
<td>AtNPF8.2 AT5G01180</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20187&lt;sup&gt;e&lt;/sup&gt;</td>
<td>scaffold_144: 44917-46935 (−)</td>
<td>0</td>
<td>609</td>
<td>12</td>
<td>AtNPF8.1 AT3G54140</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25493&lt;sup&gt;e&lt;/sup&gt;</td>
<td>scaffold_225: 74815-76805 (−)</td>
<td>0</td>
<td>609</td>
<td>12</td>
<td>AtNPF8.1 AT3G54140</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td><em>S. marinoi</em></td>
<td>29291</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AtNPF8.2 AT5G01180</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12448</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AtNPF5.2 AT5G46050</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td><em>C. cryptica</em></td>
<td>26601</td>
<td>chr83: 3655-5942 (−)</td>
<td>1</td>
<td>737</td>
<td>12</td>
<td>AtNPF6.4 AT3G21670</td>
<td>30</td>
<td>(Traller et al., 2016)</td>
</tr>
<tr>
<td></td>
<td>35672</td>
<td>chr114: 10006-12057 (+)</td>
<td>2</td>
<td>623</td>
<td>12</td>
<td>AtNPF5.7 AT3G53960</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td><em>S. robusta</em></td>
<td>245800</td>
<td>chr1142: 15208-17268 (+)</td>
<td>0</td>
<td>686</td>
<td>12</td>
<td>AtNPF6.4 AT3G21670</td>
<td>23</td>
<td>(Osuna-Cruz et al., 2020)</td>
</tr>
<tr>
<td></td>
<td>2800</td>
<td>chr3: 245400-247466 (−)</td>
<td>0</td>
<td>688</td>
<td>12</td>
<td>AtNPF8.1 AT3G54140</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>337320</td>
<td>chr2795: 9912-11528 (−)</td>
<td>0</td>
<td>538</td>
<td>11</td>
<td>AtNPF8.2 AT5G01180</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td></td>
<td>22590</td>
<td>chr35: 134056-135699 (−)</td>
<td>0</td>
<td>547</td>
<td>12</td>
<td>AtNPF4.7 AT5G62730</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td></td>
<td>193560</td>
<td>chr727: 5736-8118 (+)</td>
<td>1</td>
<td>748</td>
<td>12</td>
<td>AtNPF2.12 AT1G27080</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>262280</td>
<td>chr1319: 16210-18853 (+)</td>
<td>1</td>
<td>849</td>
<td>12</td>
<td>AtNPF6.4 AT3G21670</td>
<td>34</td>
<td></td>
</tr>
</tbody>
</table>
2.3.2. Phylogeny of diNPFs

To elucidate the molecular evolution and phylogenetic relationships among the NPF proteins, a phylogenetic tree was constructed, and all the diNPF sequences identified from available sequenced genomes, the TARA Oceans gene atlas, and the MMETSP dataset, were included in the tree. The final NPFs alignment consisted of 448 total sequences with a length of 271 aa (including gaps), and included sequences from five out the six biological kingdoms, with only Archaea being excluded. Of the 448 sequences, 20 belonged to bacteria, 6 to fungi, 344 to protists (266 of which to diatoms), 71 to plants and 7 to animals. Phylogenetic relationships between sequences are shown in the compact phylogenetic tree in the Figure 2.1 (an extended table indicating all the diNPFs is described in the Supplementary Tables S2.1 and S2.2).

Bacterial POTs were used as an outgroup to all the other sequences. Two Chlorophyceae NPFs, from Micromonas pusilla and Ostreococcus Tauris, belong to the so called Chlorophyceae Clade I, which is a sister group to all the other ingroup genes, and are close to the bacteria POTs outgroup, suggesting that these two genes derive from a relatively recent event of horizontal gene transfer (HGT). A possible alternative to HGT, explaining the proximity between bacteria POTs and Chlorophyceae Clade I, could be a Long Branch Attraction (LBA) phenomenon. Similarly, single bacterial gene from a Pelagibacteraceae bacterium (MFS transporter) is basal to the two main NPFs clades.

The two main NPFs clades (NPFs Clade I and Clade II, respectively salmon and cyan boxes in Figure 2.1), both include genes belonging to species from four biological kingdoms (Prokaryota, Protista, Plantae and Animalia). Interestingly, in the Clade I, Viridiplantae are only represented by two genes from the common moss Physcomitrella patens, while no Streptophyta and no Chlorophyceae are included in this clade. This suggest another HGT event could be involved in the close relationship between these two only bryophyte and other sequences belonging to NPFs Clade I. diNPFs Clade I (Supplementary Table S2.1), belonging to the NPFs Clade I, includes 64 sequences, of which only two are from genes expressed in the wild (TARA Oceans MATOU, Supplementary Table S2.2).

On the other hand, diNPFs Clade II belonging to the NPFs Clade II, is a sister clade to all other Protista, Plantae and Animalia NPFs, and includes 202 diatom sequences, well represented in both public genomes, TARA Oceans data and MMETSP data. In fact, of these 202 diatom sequences, 40 (out of a total of 42) are expressed in the wild. The main
subclade within NPFs Clade II includes two Dinophyceae and one Haptophyta clades, plus one sequence from *Aureococcus anophagefferens* (Pelagophyceae) and one from *Pteridomonas danica* (Dictyochophyceae). To the same clade belongs Viridiplantae, with the only exceptions previously described. Viridiplantae includes all the Chlorophyceae, Briophyta and Spermatophyta sequences included in the phylogenetic tree (Supplementary Table S2.1). No bacterial NPFs are found in the NPFs Clade II, while all fungal NPF sequences belong to this clade (Fig. 2.1).

An in-depth analysis of the evolutionary relationships highlights that diNPFs included in the Clade I diNPFs are poorly expressed in the global ocean, with only 2, out of 42, MATOU diNPFs being found in this clade (namely, MATOU-v1_19401141 belonging to the *Pseudo-nitzschia* genus and MATOU-v1_113685264 to *Fragilariopsis kerguelensis*) (Supplementary Table S2.2). Moreover, there are a number of species which appear to have sequences only in Clade I and not in Clade II (namely, *Cyclophora tenuis*, *Dactyliosolen fragilissimus*, *Entomoneis* sp., *Grammatophora oceanica*, *Leptocylindrus* genus, *Odontella* sp., *Proboscia* genus, *Striatella unipunctata*), while an even bigger number of species/genera seem to have sequences only in diatoms Clade II (e.g. *Skeletonema* genus and *Thalassiosira* genus).

Albeit the focus of the present study is on diatoms, our phylogeny elucidates evolutionary relationships also in other phytoplankton. Interestingly, no other phytoplankton group shows the same two-clade dichotomy of diatoms: Pelagophyceae and Bolidophyceae are exclusively found in the NPFs Clade I, while Dinophyceae and Haptophyta are all exclusively found in the NPFs Clade II (where Dinophyceae show two well-separated subclades). The only exception is given by the Dictyochophyceae, which are present in both NPFs Clade I and II, although with a single sequence in each clade (Supplementary Table S2.1).

A puzzling observation is again the absence of NPFs sequences in *Chaetoceros*, one of the largest genera of diatoms, with many abundant and cosmopolitan species (Li et al., 2017b). Busseni et al. (2019) had identified high-affinity NO$_3^-$ (NRT2) and NH$_4^+$ (AMT) transporters for this genus, while, with the exception of one single sequence for *Chaetoceros* sp. (CAMPEP_0176481300), we were unable to detect any NPFs for *Chaetoceros* in the same datasets, including the transcriptome of *Chaetoceros decipiens* (Amato et al., 2017), as previously said.
Diatom Clade I and Clade II NPFs are *bona fide* distinct genes, while at the present stage it would require more in-depth studies to assert whether within-clade distinct species-specific lineages correspond to different genes or to isoforms of the same gene, mostly due to the fact that many *TARA* Oceans MATOU sequences tend to be partial/incomplete. Nonetheless, the above results are in line with those reported for diatom genomes in the previous section.

**Figure 2.1.** NPFs evolutionary relationships inferred using the maximum-likelihood and Bayesian inference approaches. Numbers over the nodes represent bootstrap values. For all bootstrap values greater than 75, a posterior probability greater than 0.75 was also found. Branches were collapsed at high taxonomical levels. Salmon box indicates the NPFs Clade I, while cyan box indicates the NPFs Clade II. Red branches are used for diNPFs Clades I and II, yellow for Metazoa, green for plants and green algae, grey for bacteria. *H. sapiens*: *Homo sapiens*, *P. danica*: *Pteridomonas danica*, *A. anophagefferens*: *Aureococcus anophagefferens*, *C. wailesii*: *Coscinodiscus wailesii*, *C. elegans*: *Chaeonobdella elegans*, *P. patens*: *Physcomitrella patens*, *R. marina*: *Rhizochromulina marina*. Complete list of sequences is reported in Supplementary Table S2.1. Image from Santin et al. (2021b).

### 2.3.3. diNPFs distribution in the global ocean

*diNPF* richness was analysed through the count of distinct *diNPFs* found in the *TARA* Oceans eukaryote Unigene catalogue (Buscelli et al., 2019; Carradec et al., 2018). Generally, *diNPFs* richness (Fig. 2.2A) is high in the South Polar region and in the Coastal biome of the Southern hemisphere. Other peaks in *diNPF* richness are found in the Adriatic Sea, in the North Atlantic Sea and in the Pacific Ocean along with the Equatorial upwelling. On the other hand, the middle North Atlantic Ocean, the Western Mediterranean basin, and some stations in the Indian Ocean show a minimum in the number of distinct *diNPFs* detected.
diNPF DNA relative abundance is shown in Figure 2.2B. Data are separately presented for two size classes, 5–20 μm and 20–180 μm respectively (Busseni et al., 2019), and for two different sampling depths: surface (SUR) around 3–5 m depth and Deep Chlorophyll Maximum (DCM) tens of meters below surface. diNPFs are usually more abundant at the DCM than at the surface, and very abundant at the South Pole, with the exception of diNPFs of the 5–20 μm size class, at the DCM.

Conversely, the comparison of diNPF mRNA occurrences (Fig. 2.2C) shows more similarity between different size classes and depth than diNPF DNA. This is interesting, since low-DNA-high mRNA (and vice versa) levels in a given sampling station are the only proper way to infer patterns of up- and downregulation of genes in the TARA Oceans dataset (Caputi et al., 2019; Sordino et al., 2020). As an example, the high-DNA-high-mRNA level detected in the South Polar sampling stations (fraction 5–20 μM, SUR, and fraction 20–180 μM, DCM and SUR) is representative of basal expression of diNPFs, while low-DNA-high-mRNA level (South Pole, fraction 5–20 μM, DCM) is likely representative of few, highly expressed diNPFs. The comparison of DNA and mRNA diNPF levels is in general suggestive of highly expressed genes at the surface, and of low expression of the same genes at the DCM (with some exceptions). Generally, diNPFs seem to be highly expressed in superficial waters in the region of the Equatorial upwelling in the two size classes, suggesting light could play a role in regulating diNPF genes expression.
**Figure 2.2.** *diNPF* richness (A), expressed as the number of different sequences retrieved in different sampling stations, and *diNPF* DNA (B) and mRNA (C) relative abundance in the global ocean. Data mapped are from the TARA Oceans dataset, and come from two size classes, 5–20 μm and 20–180 μm (Busseni et al., 2019), and for two different sampling depths: surface (SUR) around 3–5 m depth and Deep Chlorophyll Maximum (DCM) tens of meters below surface. Sampling stations are coloured according to the belonging biome. While in panel A the number of sequences retrieved is presented for all size fractions and depths, in panels B and C, data are separately represented. White dots indicate TARA Oceans stations where no *diNPFs* were present. Circle size is proportional to abundances. Image from Santin et al. (2021b).

Following the same assumption above mentioned on the way to detect transcriptional variations in the oceanic samples, *diNPF* DNA and mRNA levels in the TARA Oceans sampling stations were correlated with eco-physiological variables. In most cases, *diNPF* DNA and mRNA levels show very similar trends when correlated with environmental variables (Supplementary Figs. S2.1, S2.2 and S2.3), suggesting not strong *diNPF* gene expression regulation in response to NO$_2^\cdot$, NO$_2$NO$_3^-$, PO$_4^{3-}$, Si, Fe, temperature and diel cycle.

This is also true when *diNPF* DNA and mRNA levels are correlated with NO$_2^\cdot$ and NO$_2$NO$_3^-$ environmental concentration (Supplementary Fig. S2.1). However, when *diNPF* DNA and mRNA levels are correlated with NO$_2^\cdot$ oceanic concentrations, in two cases (namely, 20–180 μm, at both SUR and DCM sampling depths), mRNA levels show trends indicative of two types of response (Fig. 2.3). In the first type (Type I), no variation in the mRNA levels is detected in increasing NO$_2^\cdot$ environmental concentration, suggesting a long-term adaptation to NO$_2^\cdot$ (Fig. 2.3). In the second type of response (Type II), mRNA levels are increasing in constant or slightly increased NO$_2^\cdot$ environmental concentration, probably indicating a short-term acclimation to NO$_2^\cdot$ concentration (Fig. 2.3).

Acclimation and adaptation are two different concepts: the first one is a short-term response, which often involves morphology, physiology, biochemistry and gene expression adjustments within a single organism, in order to improve survival in response to specific stressors; while adaptation is a trans-generational and heritable response which involves genetic (or epigenetic) modifications and maintains or increases the fitness of an organism under a given set of environmental conditions (Borowitzka, 2018). This explains why Type I response (adaptation) is represented by a horizontal trend line, with gene expression stable at increasing NO$_2^\cdot$ values, while Type II response (acclimation) is given by rapidly increasing gene expression levels in response to small changes in NO$_2^\cdot$ concentration (Fig. 2.3).
Figure 2.3. Correlation by means of Pearson’s $r$ of diNPF DNA (Metagenomics, A and B) and mRNA (Metatranscriptomics, C and D) with NO$_2^-$ levels in the global ocean for the size class 20–180 μm, at both DCM and SUR sampling depths. In both cases, DNA and mRNA levels have been correlated one to the other and in relation to NO$_2^-$ level (E and F). Red circles in the mRNA versus NO$_2^-$ scatter plots indicate Type I and Type II responses. Image from Santin et al. (2021b).

Interestingly, the two types of response are also detected when diNPF DNA and mRNA levels are correlated with sunshine duration (SSD) (Fig. 2.4). Herein, type I and type II responses in mRNA levels are detected at the DCM (both size classes) and at SUR (20–180 μm). These results suggest that the response of diNPFs to sunshine duration is rather complex, with different communities responding differently.
Figure 2.4. Correlation by means of Pearson’s $r$ of diNPF DNA (Metagenomics, A, B and C) and mRNA (Metatranscriptomics, D, E and F) with sunshine duration (SSD) levels in the global ocean for the size class 20–180 μm, at both DCM and SUR sampling depths. In both cases, DNA and mRNA levels have been correlated one to the other and in relation to SSD level (G, H and I). Red circles in the mRNA versus SSD scatter plots indicate Type I and Type II responses. Image from Santin et al. (2021b).

2.3.4. diNPFs expression data based on literature

In recent years, several gene expression analyses were performed comparing diatom cells grown in different culture conditions, and integrated studies provided a picture of the molecular, metabolic and physiological responses to different stresses. Here, a literature survey allowed to find information on diNPFs genes expression in different growth conditions, in order to understand which factors contribute to their transcriptional regulation.

The conditions on which more works focused and so for which more data were obtained, are different N sources, in particular NO$_3^-$, NO$_2^-$, NH$_4^+$, NO$_3$NH$_4$ and urea (Allen et al., 2011; Smith et al., 2019), and different NO$_3^-$ concentrations (Allen et al., 2011; Bender et al., 2014; Remmers et al., 2018; Levitan et al., 2015; Matthijs et al., 2016; Alipanah et al., 2015; Smith et al., 2019). However, in none of these conditions the expression of diNPFs was reported to be differentially regulated, with the only exception of two NPFs
of *F. cylindrus*, *Fc147192* and *Fc200740*, which show an upregulation in response to N starvation (Bender et al., 2014) (Table 2.3). These results could suggest a possible dual-affinity transport activity of these *FcNPFs*, as already reported for *A. thaliana NFP6.3*, a transceptor which is able to function as a receptor for the external NO₃⁻ concentration and, as consequence of it, is able to regulate its transporter affinity and activity (Gojon et al., 2011).

In literature, transcriptomic analyses were also reported in response to different concentrations of other nutrients, such as silicon (Sapriel et al., 2009; Shrestha et al., 2012; Ashworth et al., 2013, 2016), phosphate (Dyhrman et al., 2012; Ashworth et al., 2013) and iron (Thamatrakoln et al., 2012; Cohen et al., 2017; Mock et al., 2017). Again, no significant regulation was observed, except for *TpNPF4104*, that show a downregulation in response to low iron concentrations (Thamatrakoln et al., 2012) (Table 2.3).

Moreover, no gene regulation was observed neither in different growth stages (*P. multiseriatus*) (Boissonneault et al., 2013), nor in different mating types (*P. multistriata*) (Basu et al., 2017; Russo et al., 2018a), nor at different temperatures (*F. cylindrus*) (Mock et al., 2017) (Table 2.3).

Instead, as regards exposure to different contaminants, *Pt47148* is upregulated in response to dispersant exposure (Hook and Osborn, 2012), while cadmium does not have an effect on transcriptional regulation of *P. tricornutum NPFs* (Brembu et al., 2011). Carvalho et al. (2011) observed a downregulation in *Tp4104* gene in response to benzopyrene.

Many studies reported the effect of light in diatom genes expression: there is the evidence of a cross-talk between light and nutrient conditions, so that light-dark cycle is able to influence not only the nutrient uptake (including the most important N) but also their redistribution inside the cell (Kamp et al., 2011, 2013). However, no differential expression of the *NPFs* of *P. tricornutum* was observed, neither at different light wavelengths (Valle et al., 2014), nor at different light intensities (Nymark et al., 2009). Instead, interesting results were observed regarding gene expression in the light-dark cycle: significant downregulation was reported in the darkness for *Pt47218* (Chauton et al., 2013) and for *Tp4104* (Ashworth et al., 2013) (Table 2.3). Furthermore, in *F. cylindrus* reported works it was possible to observe that different *diNPFs* have a different expression in response to the dark condition: in darkness, *Fc204239, Fc256377,*
*Fc171976* and *Fc136520* are downregulated while *Fc147192* and *Fc200740* are upregulated (Mock et al., 2017). The larger number of *NPFs* owned by *F. cylindrus* compared to other diatoms, which instead have two on average, together with the different regulation of these *NPFs* in response to the dark, suggests that in *F. cylindrus* these genes have evolved so as to respond to different environmental conditions, promoting the adaptation of this species in such extreme and variable environments, such as the Southern Ocean (Sallée et al., 2010).

Interestingly data obtained from transcriptomic analyses in *T. pseudonana* showed upregulation of *Tp4104* at high pH (Mock et al., 2008). As already specified, POT/PTR/NPFs are a family of proton-dependent transporters, so it is plausible that a different pH, which reflects an altered transmembrane proton gradient, could influence the expression of these genes.

Table 2.3. *diNPFs* gene expression data from literature.
<table>
<thead>
<tr>
<th>Species</th>
<th>ID NPFs</th>
<th>Condition</th>
<th>Variables</th>
<th>\textit{diNPFs} expression</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{P. tricornutum}</td>
<td>\textit{Pt47148, Pt47218}</td>
<td>light/dark cycle</td>
<td>150 µE, 16:8 h photoperiod</td>
<td>47218: downregulation in dark</td>
<td>(Chauton et al., 2013)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N concentration</td>
<td>882 µM NO₃, 50 µM or no NO₃</td>
<td>no different expression</td>
<td>(Matthijs et al., 2016; Alipanah et al., 2015; Levitan et al., 2015)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N concentration and sources</td>
<td>no N, 300 µM NH₄, 300 µM NO₂, 300 µM NO₃</td>
<td>no different expression</td>
<td>(Smith et al., 2019)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Si concentration</td>
<td>350 µM SiO₂, 175 µM SiO₂</td>
<td>no different expression</td>
<td>(Sapriel et al., 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>light intensity</td>
<td>light 35 µE, light 500 µE</td>
<td>no different expression</td>
<td>(Nymark et al., 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>cadmium</td>
<td>exposure to 123 µg/L and no Cd</td>
<td>no different expression</td>
<td>(Brembu et al., 2011)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pollutants</td>
<td>WAF, CEWAF, dispersant</td>
<td>47148: upregulation in dispersant</td>
<td>(Hook and Osborn, 2012)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>light wavelength</td>
<td>100 µE white, 50 µE blue, 100 µE green, 230 µE red light</td>
<td>no different expression</td>
<td>(Valle et al., 2014)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N, P concentration</td>
<td>36 µM PO₄, no P, 882 µM NO₃, no N</td>
<td>no different expression</td>
<td>(Alipanah et al., 2018)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N concentration</td>
<td>continuous feed 0.11 gr N /day, 0.02 gr N /day</td>
<td>no different expression</td>
<td>(Remmers et al., 2018)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N concentration and sources</td>
<td>882 µM NO₃, 100 µM NO₃, 50 µM NO₃, 75 µM NH₄, 75 µM urea</td>
<td>no different expression</td>
<td>(Allen et al., 2011)</td>
</tr>
<tr>
<td>\textit{P. multistriata}</td>
<td>\textit{Pma12290, Pma7930}</td>
<td>sex</td>
<td>MT+, MT-</td>
<td>no different expression</td>
<td>(Basu et al., 2017; Russo et al., 2018a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fe concentration</td>
<td>environmental high Fe, low Fe</td>
<td>no different expression</td>
<td>(Cohen et al., 2017)</td>
</tr>
<tr>
<td>Species</td>
<td>ID NPFs</td>
<td>Condition</td>
<td>Variables</td>
<td>Different expression</td>
<td>Reference</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------------------</td>
<td>------------------------------------------------</td>
<td>------------------------------------------------</td>
<td>---------------------------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td><em>P. multiseries</em></td>
<td>Pmu190665, Pmu226109</td>
<td>N concentration</td>
<td>882 µM NO$_3$, 50 µM NO$_3$</td>
<td>no different expression</td>
<td>(Bender et al., 2014)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fe concentration</td>
<td>environmental high Fe, low Fe</td>
<td>no different expression</td>
<td>(Cohen et al., 2017)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>growth phase</td>
<td>exponential, stationary phase</td>
<td>no different expression</td>
<td>(Boissonneault et al., 2013)</td>
</tr>
<tr>
<td><em>T. pseudonana</em></td>
<td>Tp4104, Tp269333</td>
<td>nutrient concentration, light/dark cycle</td>
<td>light 125 µE, dark with 12:12 h photoperiod, 20 µM PO$_4$, 110 µM SiO$_2$, 80 µM NO$_3$</td>
<td>4104: upregulation in light vs dark</td>
<td>(Ashworth et al., 2013)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pH</td>
<td>pH 8, pH 9.4</td>
<td>4104: upregulation in high pH</td>
<td>(Mock et al., 2008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pollutants</td>
<td>exposure to mutagen benzo[a]pyrene BaP</td>
<td>both: downregulation in BaP</td>
<td>(Carvalho et al., 2011)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fe concentration</td>
<td>1 mM FeCl$_3$, 52 nM FeCl$_3$, 102 nM FeCl$_3$, 602 nM FeCl$_3$</td>
<td>4104: downregulation in low [Fe]</td>
<td>(Thamatrakoln et al., 2012; Cohen et al., 2017)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N concentration</td>
<td>882 µM NO$_3$, 50 µM NO$_3$</td>
<td>no different expression</td>
<td>(Bender et al., 2014)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P concentration</td>
<td>36 µM PO$_4$, 0.4 µM PO$_4$</td>
<td>no different expression</td>
<td>(Dyhrman et al., 2012)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Si concentration</td>
<td>200 µM Si, no Si</td>
<td>no different expression</td>
<td>(Shrestha et al., 2012)</td>
</tr>
<tr>
<td><em>F. cylindrus</em></td>
<td>Fc186175, Fc171976, Fc136520, Fc147192, Fc200740, Fc256377, Fc204239</td>
<td>Fe concentration, light/dark cycle, pCO$_2$, temperature</td>
<td>4°C, -2°C, 11°C, +1000 ppm CO$_2$, no Fe, 7 days darkness</td>
<td>204239, 256377, 171976, 136520: downregulation in darkness 147192, 200740: upregulation in darkness</td>
<td>(Mock et al., 2017)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N concentration</td>
<td>882 µM NO$_3$, 50 µM NO$_3$</td>
<td>147192, 200740: upregulation in N starvation</td>
<td>(Bender et al., 2014)</td>
</tr>
</tbody>
</table>
2.3.5. Expression patterns of NPFs in *P. tricornutum*

Previously described exploration of the expression patterns of *diNPF* genes was performed to integrate data from *TARA* Oceans meta-omics analyses and independent gene expression studies already reported in literature (Table 2.3). Then, to determine the regulation of *diNPFs* in response to continuously changing environmental factors, the gene expression pattern of the model diatom *P. tricornutum NPFs*, *PtNPF1* (ID 47148) and *PtNPF2* (ID 47218), in wild-type Pt1 cells grown under different culture conditions, was performed (Fig. 2.5).

First, to investigate whether *diNPF* genes can respond to diverse N concentrations and sources, the expression of these genes was analysed in wild-type cells grown under N starvation (50 μM NO$_3^-$) and exposed to different N sources (1 mM NH$_4$Cl, 1 mM NH$_4$NO$_3$ or 1 mM urea) with no alteration of the expression profiles of *PtNPFs* (Fig. 2.5A). This is consistent with the data obtained from other works (Allen et al., 2011; Smith et al., 2019; Levitan et al., 2015; Remmers et al., 2018), in which it is showed that N source and/or concentration do not regulate NPFs gene expression in *P. tricornutum*.

Interestingly, a regulation of the *diNPF* transcripts in dark was previously observed in different species (Mock et al., 2017; Ashworth et al., 2013; Chauton et al., 2013) (Table 2.3). Therefore, the expression of the two *PtNPFs* genes was tested in Pt1 cells collected during the dark period (12 hours of dark), observing a clear decrease of the *PtNPF2* mRNAs, while *PtNPF1* is only slightly less transcribed (Fig. 2.5A), suggesting a diurnal regulation of *PtNPF2*.

Then, a pH effect in regulating the expression of the *PtNPF* genes, and in particular of *PtNPF2*, was observed. As reported in Figure 2.5A, *PtNPF2* shows an upregulation at lower pH compared to the control condition, while at alkaline pH a *PtNPF2* downregulation can be observed, suggesting that this transporter could be pH sensitive. This result is consistent with the results reported in Mock et al. (2008), in which a pH dependent regulation of one of the two *TpNPFs* genes was observed, and with different works on bacterial POTs and plant NPFs (Solcan et al., 2012; Newstead, 2017), in which the ExxER structural motif was demonstrated to be involved in H$^+$ transport, supporting our hypothesis of a H$^+$-dependent transport of diNPFs.

Ten ecotypes of *P. tricornutum* from different ecological niches, from sub-polar to tropical latitudes, have been identified and characterised (De Martino et al., 2007; Rastogi
et al., 2020). PtNPFs expression levels were analysed in ecotype Pt4, known to have adapted to high latitudes environmental conditions (De Martino et al., 2007). It possesses a very high proportion of specific Single Nucleotide Polymorphisms (SNPs ~35%) and Insertions/Deletions (INDELs ~75%) (Rastogi et al., 2020). Furthermore, the Pt4 ecotype displayed systematically reduced NPQ levels with respect to other strains, suggesting it is a natural NPQ variant, most probably reflecting a constitutive adaptation to its environment (Bailleul et al., 2010). The identification of a natural NPQ variant that displays reduced expression of the gene encoding for the Light-Harvesting Complex Protein 1 of the PSII (Lhcx1), suggested to check if there was any particular expression profile also for the PtNPFs. Again, no differential expression has been observed between the two ecotypes Pt1 and Pt4 (Fig. 2.5B).

![Figure 2.5](image.png)

**Figure 2.5.** Expression profile analysis of PtNPFs genes as assessed by qPCR in A) different N concentrations, in different N sources, in light or dark condition, at different pHs, and B) in different ecotypes (Pt1 as control and Pt4). Experiments were performed separately, using as control *P. tricornutum* wild-type Pt1, grown in 882 μM NO₃⁻ at pH 8, collected during the light phase. Blue bars: logFC (fold-changes) for PtNPF1; red bars: logFC for PtNPF2. Black bars represent standard deviations. Light-blue dotted lines indicate significance thresholds. Image adapted from Santin et al. (2021b).

### 2.3.6. Upstream regulation of *diNPF* genes

Transcriptional regulation is a dynamic and complex process that in eukaryotic organisms requires the recognition of specific DNA sequences (*cis*-acting elements) by Transcription Factors (TFs) (*trans*-acting factors), to favour recruitment of the RNA polymerase for mRNA synthesis initiation (Russo et al., 2015). A catalogue of all
predicted TFs of \textit{P. tricornutum} and \textit{T. pseudonana} was published (Rayko et al., 2010), while recent reports have identified new diatom TFs (Huysman et al., 2013; Matthijs et al., 2017). Nonetheless, transcriptional control mechanisms currently remain poorly understood in diatoms and over 90% of TFs still have no function assigned (Matthijs et al., 2017).

Despite this, due to the observed transcriptional responses of \textit{diNPF} genes at different light and pH conditions, and with the aim of identifying possible regulatory motifs involved in the regulation of the \textit{diNPFs} transcription, a search for known and potentially novel regulatory motifs in the 5′-flanking regions of \textit{diNPF} genes was done, using MEME Suite. MEME Suite web server provides a portal for online discovery and analysis of sequence motifs representing features, such as DNA binding sites and protein interaction domains, and for comparison of TF motifs (including those discovered using MEME) with motifs already known in many popular databases (Bailey et al., 2009).

So, putative regulatory regions of the \textit{diNPF} genes set were used to identify motifs that could function as a TF-binding site (TFBS). For the species in which the genomes were sequenced and public, it was possible to identify the first intergenic region of about 500-1000 bp upstream of \textit{diNPF} genes, which was the input for MEME. In particular, the upstream sequences of the \textit{NPF} genes of \textit{P. tricornutum}, \textit{P. multistriata}, \textit{T. pseudonana} and \textit{F. cylindrus} were obtained. In detail, three motifs were identified in all the sequences analysed, and these motifs were compared using TOMTOM to the JASPAR CORE database, a curated set of eukaryotic TFBS, to determine the class of TFs that might bind and regulate these sites. This returned 5 motifs, that bind TFs found in other organisms as plants or fungi through a zinc-finger domain.

Moreover, other two motifs were identified in many but not all upstream NPF sequences provided. In particular, the most represented of these (the red one in Figure 2.6), returned 11 motifs when compared with a Eukaryotic motifs database using TOMTOM.
Figure 2.6. Potential cis-regulatory motifs in the diNPF 5'-flanking regions with their sequence localisation, for *P. tricornutum*, *P. multistriata*, *T. pseudonana* and *F. cylindrus*. The discovered motifs are indicated on each 5'-flanking sequence of diNPF genes, highlighted in different bar colours and with a bar height equal to their significance. The sequences have different lengths because they correspond to the entire 5'-intergenic region flanking the diNPF genes. Image from Santin et al. (2021b).

Most of these motifs bind plant TFs involved in the regulation of gene expression by environmental or stress factors. In particular, 7 plant TFs models belong mainly to Ethylene-Responsive Transcription Factors (ERF) class, which regulate transcription in response to hormone levels like ethylene, abscisic acid or IAA (indole-3-acetic acid).

Finally, FIMO was used to verify the presence of two NO$_3^-$ regulated motifs HNS_A and HNS_B, previously identified by Smith et al. (2019) in upstream diNPF sequences. These motifs were identified only in *Fc136520*, *Pt47148* and *Pt47218*, (*p*-value > 10$^{-5}$) (Table 2.4).

Table 2.4. NO$_3^-$ regulator motifs in upstream diNPF sequences.

<table>
<thead>
<tr>
<th>Motif ID</th>
<th>Sequence name</th>
<th>Position</th>
<th>p-value</th>
<th>q-value</th>
<th>Matched sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>HNS_A</td>
<td>Fc136520_prom</td>
<td>386-408 (-)</td>
<td>1.02e-05</td>
<td>0.214</td>
<td>CATCCGCCGGAAG</td>
</tr>
<tr>
<td>HNS_A</td>
<td>Pt47148_prom</td>
<td>403-415 (+)</td>
<td>6.35e-05</td>
<td>0.666</td>
<td>GCGGCGCGCGCAG</td>
</tr>
<tr>
<td>HNS_B</td>
<td>Pt47218_prom</td>
<td>624-636 (+)</td>
<td>3.17e-05</td>
<td>0.653</td>
<td>GATTTGACAGTGAAG</td>
</tr>
</tbody>
</table>
2.3.7. Structural modelling of diNPFs

In order to obtain structural models of diNPFs, NPF amino acid sequences from all diatoms public genomes (Table 2.2) were aligned. Crystal structures of the *A. thaliana* NPF6.3 (pdb: 4oh3) (Sun et al., 2014) and of the *Shewanella oneidensis* peptide transporter PepTso (pdb: 4uvm) (Fowler et al., 2015) (see details about pdb used in Table 2.5) were used as template structures. The structural models for PtNPF1 and PtNPF2 from *P. tricornutum* are here reported as representatives of Clade II and Clade I diNPFs, respectively, in Figure 2.7. Structure homology models were obtained using the SWISS-MODEL workspace (Arnold et al., 2006; Biasini et al., 2014).

Complete diNPF sequences belonging to the two Clades are predicted to fold into 12 TMHs organised in two bundles, the N- and the C-terminal domains (Fig. 2.7). However, structural models obtained together with multiple sequence alignment reveal differences in the length and position of the loops between TMHs.

![Separated and overlapped structural models of *P. tricornutum* PtNPF1 (Clade II) and PtNPF2 (Clade I). A) The PtNPF1 model (N-terminal in blue, C-terminal in magenta and lateral helix in yellow) was obtained using the crystal structure of the *A. thaliana* NPF6.3 (pdb: 4oh3) as template (Sun et al., 2014). B) The PtNPF2 model was obtained using both the AtNPF6.3 and the *S. oneidensis* peptide transporter PepTso (pdb: 4uvm) (Fowler et al., 2015) as template structures. C) The two structural models overlapped to highlight their structural differences between diNPF belonging to two different clades. Image from Santin et al. (2021b).](image-url)
First, all Clade I diNPFs present an additional amino acid sequence of variable length between TMH4 and 5, while none of the Clade II diNPFs has it in this position (Table 2.5 and Fig. 2.7). Second, both Clade I and II diNPFs contain an insert sequence between TMH6 and TMH7 (Table 2.5 and Fig. 2.7). In Clade II diNPFs this insert is predicted to fold as a later helix like in plant NPFs, while in Clade I diNPFs, it is predicted to fold into a later helix and into two TMHs as observed in the crystal structure of bacterial POTs. There is not a known role for these additional TMHs, but since they are absent in fungal, mammalian and plant transporters, it has been suggested they contribute to stability or folding more than to the transport mechanism (Newstead, 2015).

Table 2.5. Properties of PtNPFs and templates used to build their structural homology models. Pdb: Protein Data Bank ID.

<table>
<thead>
<tr>
<th>Clade</th>
<th>Protein</th>
<th>Length (AA)</th>
<th>Lateral helix</th>
<th>Pdb of templates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clade I</td>
<td>PtNPF2</td>
<td>650</td>
<td>Long loop between TMH4–5 Later helix + 2 TMH between TMH6–7</td>
<td>4uvvm</td>
</tr>
<tr>
<td>Clade II</td>
<td>PtNPF1</td>
<td>775</td>
<td>Later helix</td>
<td>4oh3</td>
</tr>
</tbody>
</table>

Sequence alignment and structural models of diNPFs allowed to identify key amino acids involved in proton or substrate binding and transport (Table 2.5 and Fig. 2.8). First, the ExxER motif was searched: it has an important role in coupling H⁺ binding to peptide or NO₃⁻ transport. In the crystal structures from bacterial POTs and one plant NPF, the motif is located on TMH1 with the chargeable amino acids available for H⁺ binding in the substrate cavity. In diNPFs, the chargeable amino acids of the ExxER motif are strictly conserved and our structural models confirm that the chargeable residues are correctly oriented in the access cavity (Table 2.5 and Figs. 2.8 and 2.9).
Salt bridges were identified in POTs and NPFs and were shown to be implicated in orchestrating conformational changes of the transporters and contributing to the alternating-access mechanism (Gojon et al., 2011; Newstead, 2017; Jacquot et al., 2017). One salt bridge between residues on TMH4 and TMH10 forms in the outward-open conformation (Newstead, 2015, 2017). These residues are conserved in all diNPFs (Fig. 2.9 and Table 2.6). A second salt bridge between TMH1 and TMH7 that is predicted to form in the inward open conformation in some bacterial POTs but not in plant NPFs (with the exception of two NPFs from a moss) is conserved in all Clade I diNPFs, but not in Clade II. Interestingly, Clade I diNPFs align with the two moss NPFs with residues that can form the TMH1–TMH7 salt bridge (Longo et al., 2018).

Then, residues involved in substrate binding were identified. The NO$_3^-$ binding site in the plant AtNPF6.3 crystal structure features a protonated histidine, His356 which forms an electrostatic interaction with the NO$_3^-$ (Parker and Newstead, 2014; Sun et al., 2014). Mutating His356 to alanine completely abolishes the NO$_3^-$ transport function of AtNPF6.3 (Parker and Newstead, 2014). But some plant orthologues of AtNPF6.3 have a tyrosine in this same position. In diatoms, NPFs have a conserved tyrosine in the corresponding position (Figs. 2.8 and 2.9 and Table 2.6). Another tyrosine (or...
phenylalanine) from TMH1 is located in close proximity and may contribute to creating a hydrophobic pocket for the substrate (Figs. 2.8 and 2.9 and Table 2.6).

Lastly, threonine 101 in AtNPF6.3 was considered, and the corresponding residue was searched in the diNPF sequences. The phosphorylation of Thr101 is a key mechanism in AtNPF6.3 as it allows to switch from low- to high-affinity state, and so to be called dual-affinity NO$_3^-$ transporter (Sun et al., 2014). Furthermore, dephosphorylation of Thr101 is required for the low-affinity sensor function of AtNPF6.3 in the same manner as found in many common cell surface receptors, giving the possibility to work with a dual NO$_3^-$ transport/signalling function as transceptor (Gojon et al., 2011; Sun et al., 2014). Thr101 is conserved in many, but not all, of the diNPF sequences. For example, in *P. tricornutum, P. multistriata, P. multiseries, F. cylindrus, S. marinoi* and *S. robusta* this residue is present at least in one NPF sequence, while in other NPF sequences a different amino acid was found in correspondence of the Thr101 of AtNPF6.3, in particular another polar amino acid such as serine, or non-polar amino acids such as methionine, alanine, valine or glycine. While in *T. pseudonana, F. solaris* and *C. cryptica* NPF sequences this residue is highly conserved, in *T. oceanica* both NPFs show an asparagine in this position (Table 2.6).

**Table 2.6.** Conserved residues involved in proton or substrate binding and transport.
<table>
<thead>
<tr>
<th>Diatom</th>
<th>Protein ID</th>
<th>ExxER motif</th>
<th>Salt Bridge (TMH4-10)</th>
<th>Nitrate Bond</th>
<th>H⁺ Bond (TMH7)</th>
<th>Thr 101</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>P. tricornutum</em></td>
<td>47148</td>
<td>E194 L195 L196 E197 R198</td>
<td>K310, E549</td>
<td>Y202, Y439</td>
<td>E549</td>
<td>S260</td>
</tr>
<tr>
<td></td>
<td>47218</td>
<td>E61 T62 G63 E64 R65</td>
<td>K169, E548</td>
<td>F69, Y439</td>
<td>E548</td>
<td>T120</td>
</tr>
<tr>
<td><em>T. pseudonana</em></td>
<td>4104</td>
<td>E190 M191 L192 E193 R194</td>
<td>K306, E552</td>
<td>Y198, Y443</td>
<td>E552</td>
<td>T256</td>
</tr>
<tr>
<td></td>
<td>269333</td>
<td>E64 L65 L66 E67 R68</td>
<td>K182, E414</td>
<td>Y72, Y304</td>
<td>E414</td>
<td>T130</td>
</tr>
<tr>
<td><em>T. oceanica</em></td>
<td>32021</td>
<td>E63 L64 L65 E66 R67</td>
<td>K179, E425</td>
<td>Y71, Y319</td>
<td>E425</td>
<td>N129</td>
</tr>
<tr>
<td></td>
<td>14591</td>
<td>E59 L60 L61 E62 R63</td>
<td>K175, -</td>
<td>Y67, -</td>
<td>-</td>
<td>N125</td>
</tr>
<tr>
<td><em>P. multiseries</em></td>
<td>190665</td>
<td>E13 T14 A15 E16 R17</td>
<td>K121, E520</td>
<td>F21, Y409</td>
<td>E520</td>
<td>T73</td>
</tr>
<tr>
<td></td>
<td>226109</td>
<td>E12 L13 L14 E15 R16</td>
<td>K138, E355</td>
<td>Y20, Y245</td>
<td>E355</td>
<td>A87</td>
</tr>
<tr>
<td><em>P. multistriata</em></td>
<td>7930</td>
<td>E116 L117 L118 E119 R120</td>
<td>K242, E519</td>
<td>Y124, Y409</td>
<td>E519</td>
<td>A191</td>
</tr>
<tr>
<td></td>
<td>12290</td>
<td>E72 T73 A74 E75 R76</td>
<td>K194, D544</td>
<td>F80, Y439</td>
<td>E544</td>
<td>T132</td>
</tr>
<tr>
<td><em>F. cylindrus</em></td>
<td>186175⁹</td>
<td>E29 T30 A31 E32 R33</td>
<td>K141, E455</td>
<td>F37, Y350</td>
<td>E455</td>
<td>T89</td>
</tr>
<tr>
<td></td>
<td>204239⁹</td>
<td>E30 T31 A32 E33 R34</td>
<td>K142, E456</td>
<td>F38, Y351</td>
<td>E456</td>
<td>T90</td>
</tr>
<tr>
<td></td>
<td>136520⁵</td>
<td>E65 L66 L67 E68 R69</td>
<td>K191, E413</td>
<td>Y73, Y303</td>
<td>E413</td>
<td>A140</td>
</tr>
<tr>
<td></td>
<td>256377⁷</td>
<td>-</td>
<td>- , E163</td>
<td>- , Y53</td>
<td>E163</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>171976</td>
<td>E23 L24 F25 E26 R27</td>
<td>K139, E376</td>
<td>Y31, Y266</td>
<td>E376</td>
<td>S89</td>
</tr>
<tr>
<td></td>
<td>147192</td>
<td>E54 F55 F56 E57 R58</td>
<td>K177, E417</td>
<td>Y62, Y308</td>
<td>E417</td>
<td>G127</td>
</tr>
<tr>
<td></td>
<td>200740</td>
<td>E66 A67 M68 E69 R70</td>
<td>K182, E434</td>
<td>Y74, Y325</td>
<td>E434</td>
<td>S132</td>
</tr>
<tr>
<td>Diatom</td>
<td>Protein ID</td>
<td>ExxER motif</td>
<td>Salt Bridge (TMH4-10)</td>
<td>Nitrate Bond</td>
<td>H⁺ Bond (TMH7)</td>
<td>Thr 101</td>
</tr>
<tr>
<td>---------------------</td>
<td>------------</td>
<td>--------------</td>
<td>-----------------------</td>
<td>--------------</td>
<td>----------------</td>
<td>---------</td>
</tr>
<tr>
<td><strong>F. solaris</strong></td>
<td>15278&lt;sup&gt;c&lt;/sup&gt;</td>
<td>E28 T29 G30 E31 R32</td>
<td>K134, E437</td>
<td>F36, Y332</td>
<td>E347</td>
<td>T88</td>
</tr>
<tr>
<td></td>
<td>26459&lt;sup&gt;c&lt;/sup&gt;</td>
<td>E37 T38 G39 E40 R41</td>
<td>K143, E446</td>
<td>F45, Y341</td>
<td>E446</td>
<td>T97</td>
</tr>
<tr>
<td></td>
<td>17535&lt;sup&gt;d&lt;/sup&gt;</td>
<td>E89 L90 L91 E92 R93</td>
<td>K205, E443</td>
<td>Y97, Y333</td>
<td>E443</td>
<td>T155</td>
</tr>
<tr>
<td></td>
<td>17972&lt;sup&gt;d&lt;/sup&gt;</td>
<td>E89 L90 L91 E92 R93</td>
<td>K205, E443</td>
<td>Y97, Y333</td>
<td>E443</td>
<td>T155</td>
</tr>
<tr>
<td></td>
<td>20187&lt;sup&gt;e&lt;/sup&gt;</td>
<td>E61 W62 F63 E64 R65</td>
<td>K177, E402</td>
<td>Y69, Y292</td>
<td>E402</td>
<td>T127</td>
</tr>
<tr>
<td></td>
<td>25493&lt;sup&gt;e&lt;/sup&gt;</td>
<td>E61 W62 F63 E64 R65</td>
<td>K177, E402</td>
<td>Y69, Y292</td>
<td>E402</td>
<td>T127</td>
</tr>
<tr>
<td><strong>S. marinoi</strong></td>
<td>29291</td>
<td>R101 L102 V103 E104 R105</td>
<td>D217, F480</td>
<td>Y109, Y371</td>
<td>F480</td>
<td>M166</td>
</tr>
<tr>
<td></td>
<td>12448</td>
<td>E12 M13 L14 E15 R16</td>
<td>K129, E393</td>
<td>Y20, Y284</td>
<td>E393</td>
<td>T78</td>
</tr>
<tr>
<td><strong>C. cryptica</strong></td>
<td>26601</td>
<td>E151 M152 L153 E154 R155</td>
<td>K267, E513</td>
<td>Y159, Y404</td>
<td>E513</td>
<td>T217</td>
</tr>
<tr>
<td></td>
<td>35672</td>
<td>E98 L99 L100 E101 R102</td>
<td>K216, E450</td>
<td>Y106, Y340</td>
<td>E450</td>
<td>T164</td>
</tr>
<tr>
<td><strong>S. robusta</strong></td>
<td>245800</td>
<td>E104 L105 L106 E107 R108</td>
<td>K220, E455</td>
<td>Y112, Y346</td>
<td>E455</td>
<td>G170</td>
</tr>
<tr>
<td></td>
<td>2800</td>
<td>E131 L132 L133 E134 R135</td>
<td>K247, E490</td>
<td>Y139, Y375</td>
<td>E490</td>
<td>S197</td>
</tr>
<tr>
<td></td>
<td>337320</td>
<td>-</td>
<td>K98, E348</td>
<td>- , Y239</td>
<td>E348</td>
<td>V48</td>
</tr>
<tr>
<td></td>
<td>22590</td>
<td>E21 F22 S23 E24 R25</td>
<td>K128, E440</td>
<td>Y29, Y332</td>
<td>E587</td>
<td>T81</td>
</tr>
<tr>
<td></td>
<td>193560</td>
<td>E92 T93 A94 E95 R96</td>
<td>K214, E587</td>
<td>F100, Y479</td>
<td>E440</td>
<td>V151</td>
</tr>
<tr>
<td></td>
<td>262280</td>
<td>E106 A107 A108 E109 R110</td>
<td>K217, E654</td>
<td>Y114, Y545</td>
<td>E654</td>
<td>T166</td>
</tr>
</tbody>
</table>
A *Shewanella oneidensis* PepT1 (pdb: 4uvm)

- K127-E119 = TMH4-TMH10 salt bridge - outward open conformation
- R32-D316 = TMH1-TMH7 salt bridge - inward open conformation

B *Phaeodactylum tricornutum* NPF2 (Clade I)

- K109-E348 = TMH4-TMH10 salt bridge - outward open conformation
- R32-D440 = TMH1-TMH7 salt bridge - inward open conformation

C *Arabidopsis thaliana* NPF6.3 (pdb: 4oh3)

- K164-E176 = TMH4-TMH10 salt bridge - outward open conformation

D *Phaeodactylum tricornutum* NPF1 (Clade II)

- K310-E349 = TMH4-TMH10 salt bridge - outward open conformation
Figure 2.9. Transmembrane topology of bacterial, plant and diatom POTs/NPFs. Topology plots showing the 12 transmembrane portions of A) *Shewanella oneidensis* PepT1, based on its crystal structure (pdb: 4uvm) (Fowler et al., 2015); B) *P. tricornutum* NPF2, based on its structural model; C) *Arabidopsis thaliana* NPT6.3, based on its crystal structure (pdb: 4oh3) (Sun et al., 2014); D) *P. tricornutum* NPF1, based on its structural model. ExxER motif amino acids are indicated by red squares; arginine/lysine in blue; glutamic acid/aspartic acid in red. Residues forming a salt bridge between TMH1 and TMH7 or TMH4 and TMH10 are enclosed within a yellow or green circle, respectively. Figure created with the Protter web application (http://wlab.ethz.ch/protter). Image from Santin et al. (2021b).

2.3.8. Predicted subcellular localisation

Subcellular localisation of the proteins is a key information for studying both evolution and function of proteins. Even if caution is required as the predicted subcellular localisation of the diNPFs is not experimentally validated, first working hypothesis can be drawn thanks to it.

The subcellular localisation of 325 amino acid diNPFs sequences, obtained from the sequenced genomes and from *TARA* and MMETSP datasets, was predicted by exploiting the LocTree3 software, that predicts the localisation also via homology-based inference between proteins of known localisation (Goldberg et al., 2014). In particular, it can predict the subcellular compartment of a protein by mimicking the mechanism of cellular sorting and exploiting a variety of sequence and predicted structural features in its input (Nair and Rost, 2005). The reliability of the LocTree3 software was previously tested by confirming the sub-localisation of different plant sequences whose localisation has been experimentally verified by Busseni et al. (2019), and was also exploited to predict the subcellular localisation of diatom NH$_4^+$ (diAMT) and high-affinity NO$_3^-$ (diNTR2) transporters (Busseni et al., 2019).

66% of these sequences are predicted to be plasma membrane proteins (Table 2.7), and this can be explained by the fact that NPFs are transporters and so transmembrane proteins which can play a role in the transport of different molecules from the outside into the cells. These predictions are based on sequence homology with i) plant NPFs, mainly *A. thaliana* PTR1 (or NPF8.1) and PTR5 (or NPF8.2), that are involved di- and tripeptides transport; ii) a *C. elegans* peptide transporter PEPT2, responsible of H$^+$-dependent uptake of di- or tripeptides; and iii) two bacteria peptide permeases, *E. coli* DtpD (YbgH) and DtpB (YhiP), which are H$^+$-dependent transporters of di- and tripeptides (Table 2.7).

Interestingly, a significant percentage of sequences was predicted to the vacuole membrane (25%), by homology with PTR2 of *A. thaliana*, also called NPF8.3, a vacuole
protein that mediates the transport of di- and tripeptides and which can recognise a variety of different amino acid combinations (Table 2.7). It is a H⁺-coupled, voltage-dependent acid-activated transporter (Corratgé-Faillie and Lacombe, 2017). Furthermore, since bacterial POTs are essentially oligopeptide transporters, while plant NPFs are capable to carry different substrates, including NO₃⁻ and other molecules, it is interesting that all diNPFs sequences predicted to the vacuole fall within the diNPFs Clade II, suggesting reallocation processes between different intracellular compartments. Finally, 8% of sequences were predicted to the mitochondrion membrane (Table 2.7).

As previously said, these bioinformatic predictions still requires functional characterisation, but these new and interesting results could reinforce the hypothesis that diNPFs are not only involved in molecules uptake from the external environment, but rather in their reallocation inside the cell, probably promoting intracellular NO₃⁻ storage and communication between the different organelles.

**Table 2.7.** Predicted subcellular localisation of diNPFs, including sequence number and corresponding percentage that are predicted to a specific subcellular localisation, and homologous Eukaryota and bacteria proteins that drove these predictions. * indicates that LocTree3 does not find a homologous sequence whose exact location is known, but based on the homology with different sequences, the algorithm is however able to reconstruct the path that leads a protein to locate in a given compartment.

<table>
<thead>
<tr>
<th>Localisation</th>
<th>Nº Sequences</th>
<th>% Sequences</th>
<th>Homologous Proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasma membrane</td>
<td>216 (52 Clade I and 164 Clade II)</td>
<td>66%</td>
<td>PTR1 (NPF8.1)–PTR5 (NPF8.2) <em>Arabidopsis thaliana</em> PEPT2 <em>Caenorhabditis elegans</em> (peptide transporter) DtpD–DtpB <em>Escherichia coli</em> (peptide permease)</td>
</tr>
<tr>
<td>Mitochondrion membrane</td>
<td>26 (7 Clade I and 19 Clade II)</td>
<td>8%</td>
<td>reconstructed path by software *</td>
</tr>
<tr>
<td>Vacuole membrane</td>
<td>80 (only Clade II)</td>
<td>25%</td>
<td>PTR2 (NPF8.3) <em>Arabidopsis thaliana</em> (Protein NRT1/PTR Family)</td>
</tr>
<tr>
<td>Secreted</td>
<td>2 (only Clade II)</td>
<td>1%</td>
<td>reconstructed path by software *</td>
</tr>
<tr>
<td>Endoplasmic reticulum</td>
<td>1 (Clade II)</td>
<td>0%</td>
<td>reconstructed path by software *</td>
</tr>
<tr>
<td>Total</td>
<td>325</td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>
2.4. Discussion

Diatoms are known to cope with conditions that are not favourable for other phytoplankton, in particular they have the capability to deal with fluctuations and often low availability of nutrients, especially low iron and N (Alipanah et al., 2015; Lampe et al., 2018). Several interesting studies are available on the N metabolism in diatoms (Allen et al., 2011), from diatom response to environmental N changes (Scarsini et al., 2022; Alipanah et al., 2015), to intracellular N fluxes between different organelles (Smith et al., 2019) and to the functional studies on the Nitrate Reductase (NR), an enzyme involved in N assimilation which represent one of the best examples of genes functionally characterised in diatoms (McCarthy et al., 2017). These studies revealed that diatoms possess unique features of N metabolism. However, many questions are still open, and the role of many components known to have important functions in other systems are still undefined.

To transport NO\textsubscript{3}\textsuperscript{-}, diatoms mostly rely on high-affinity transporters, NRT2s (Rogato et al., 2015), coherently with the need to deal with very low NO\textsubscript{3}\textsuperscript{-} concentrations in seawater, where NO\textsubscript{3}\textsuperscript{-} is present in the order of micromolar. However, diatoms own also low-affinity transporters, which has been shown in other organisms to be active at higher NO\textsubscript{3}\textsuperscript{-} concentrations, usually not reached in diatom marine environment. While higher plants harbour a very big number of NPFs, till 53 in \textit{A. thaliana}, diatoms mostly contain two NPFs. In this Chapter, an extensive survey of NPFs conservation in diatom genomes, transcriptomes, metagenomes and metatranscriptomes data indicates that the two NPFs homologues are broadly retained. Whether diNPFs role is to transport NO\textsubscript{3}\textsuperscript{-} or another substrate, remains to be experimentally demonstrated. Here, through obtained results, different lines of evidence which provide working hypotheses can be described.

The evolutionary analysis shows two distinct diNPFs clades, each containing one of the two homologues for most of the species analysed, with some interesting exceptions whereby some species had two genes clustered together in a single clade and no gene in the other clade. The specificity of the evolutionary processes in diNPFs is even more highlighted by the unicity of their two-clade dichotomy among the phytoplankton. This, together with the absence of NPFs in the genus \textit{Chaetoceros}, is suggestive of an ongoing process of evolution of this family in diatoms. As to why \textit{Chaetoceros} represents an exception, the NPFs absence may be due to the presence of alternative genes with conserved function or to different adaptations for N transport. It is to note that in
environments with fluctuating N, *Chaetoceros* can preferentially transport NH$_4^+$ derived from N-fixing bacteria in the phycosphere (Olofsson et al., 2019), so a possible evolutionary adaptation related to this unique feature can be possible.

The NPFs evolutionary relationships are complex, and at the present stage it is not clear whether these are resulting from events of HGT, from an even more complex pattern of specific gene loss or from a mixture of the two. It is, however, of note that Clade I and Clade II NPFs differ in kingdom specific inclusion: in fact, Bacteria are present only in Clade I, while Fungi only in Clade II; Viridiplantae are represented only by two Bryophyta sequences in Clade I (probably due to a HGT event), while Chlorophyceae and especially Streptophyta are present in Clade II.

The structural analysis confirms these phylogenetic observations, revealing that Clade I diNPFs have structural features that are found in bacterial POTs, but not in higher plants, while Clade II diNPFs are structurally closer to plant NPFs. This indicates that the NPFs repertoire in diatoms is made up of genes with a specific evolutionary history, which may be reflected in functional specialisation (Caputi et al., 2019).

Moreover, several gene expression analyses, performed in the recent years and comparing diatom cells grown in different culture conditions, were integrated. Among the conditions most widely studied, in different N sources (Allen et al., 2011; Smith et al., 2019), and in different NO$_3^-$ concentrations (Alipanah et al., 2015; Bender et al., 2014; Levitan et al., 2015; Remmers et al., 2018; Matthijs et al., 2017), the expression of diNPFs do not show a different regulation, with the only exception of two *F. cylindrus* NPFs (Bender et al., 2014) (Table 2.3).

As regards to light, puzzling results were obtained: despite no differential expression has been reported in the literature, neither at different light wavelengths (Valle et al., 2014) nor at different light intensities (Nymark et al., 2009), meta-omics analyses of diNPFs expression profile in the ocean and comparison of DNA and mRNA diNPFs levels are in general suggestive of highly expressed genes at surface and equatorial regions, where light intensities are particularly high (Figure 2.2). Moreover, interesting results were observed gene expression in the light-dark cycle: significant downregulation was reported in the darkness for *PtNPF2* (Chauton et al., 2013), confirmed by actual *PtNPF* genes expression analyses (Fig. 2.5 and Table 2.3). It is of note that analyses of the diNPFs expression profile in the ocean show two response patterns (one indicative of long-term adaptation, the other of short-term acclimatation) to environmental SSD in specific
sampling stations. Furthermore, in *F. cylindrus* different NPFs have a different expression in response to the dark condition (Mock et al., 2017). The different regulation of the *FcNPFs* in response to the dark suggests that in *F. cylindrus* these genes have evolved differently to respond to different environmental conditions, promoting the adaptation of this species in the Southern Ocean extreme and variable environments, where daily and seasonal variations, thickness of the pack and deep mixing result in low average of light intensities so that diatoms experience considerable periods of light limitation (Sallée et al., 2010).

Intriguingly, a pH effect was also highlighted in regulating the expression of the *PtNPF* genes, and in particular of *PtNPF2* (Fig. 2.5). This result is consistent with the results reported in Mock et al. (2008) (Table 2.3), in which transcriptomic analyses in *T. pseudonana* show upregulation of *Tp4104* at high pH.

In bacterial POTs and most plant NPFs, the residues responsible for H⁺ binding and transport are two highly polar glutamate residues typical of the ExxER motif located on TMH1 and for some POTs a glutamic acid found on TMH7 (Longo et al., 2018; Newstead, 2017; Solcan et al., 2012). As already seen, residues in the ExxER motif are conserved in diNPFs, suggesting a sensitivity to the pH that can trigger the conformational change, confirming that the mechanism of substrate transport in coordination with H⁺ transport is conserved in diNPFs. As an alternative, the pH-dependent regulation of expression could also be linked to a NO₃⁻-mediated mechanism for alleviating acidic pH toxicity as recently reported in higher plants (Fang et al., 2016).

Structural information, and in particular conserved key residues identified in diNPF sequences, allowed to hypothesise a model of diNPFs functioning, which would share the same alternate access mechanism as higher plant NPFs and bacterial POTs (Newstead, 2017; Solcan et al., 2012) (Fig. 2.10): i) the protein is in an open conformation towards the extracellular space stabilised by a salt bridge between TMH4 and TMH10, present in all diNPF sequences; ii) once the substrate and two H⁺ bind to the transporter through the specific ExxER motif, the protein moves to an occluded conformation; iii) the protein opens towards the intracellular space so H⁺ and substrate are released in the cytoplasm. At this point a new salt bridge may form between TMH1 and 7 in the inward open conformation to stabilise it. However, residues that form this bridge are only conserved in some bacterial POTs and in Clade I diNPFs (Fig. 2.9). Clade II diNPFs and all plant
NPFs do not have residues that form this bridge, with the only exception of the two NPFs from the moss *P. patens* that cluster with Clade I diNPFs (Figs. 2.9 and 2.10).

**Figure 2.10.** Proposed model for diNPFs functioning, with alternating-access mechanism. Dashed red lines represent salt bridges: the TMH4–TMH10 salt bridge is present in both diNPFs clades, while TMH1–TMH7 salt bridge is only present in Clade I diNPFs. Small yellow circles represent protons, while large blue circles represent a substrate, which could be NO$_3^-$ or other molecules. Image from Santin et al. (2021b).

While in organisms like bacteria and higher plants the subcellular localisation of NPFs has been widely studied and experimentally validated, for marine unicellular phototrophs, whose environment is characterised by fluctuating concentrations of their substrates, no data is available. As first approach, bioinformatic prediction of subcellular localisation of diNPFs was done, observing most of the sequences predicted to the plasma membrane, as in the case of most higher plant NPFs, which play roles in NO$_3^-$ loading and unloading in cells of different tissues (Wang et al., 2020), and bacterial POTs, which are mainly involved in the uptake from the outside (Parker et al., 2017).

Considering that the concentration of NO$_3^-$ in the ocean is in the order of micromolar, the analysis of *TARA* Oceans database and the gene expression profiles suggest that the expression of these genes is not strongly regulated by external NO$_3^-$ availability, although a weak correlation with NO$_2^-$ levels was observed. Moreover, the predicted protein structures cannot be used to determine the substrate specificities of NPF transporter, and the transported substrate can be only defined with an accurate biochemical characterisation of the proteins. At this point, it is only possible to speculate that diNPFs could have evolved the capability to transport substrates different than NO$_3^-$, including hormones and peptides, but more in-depth analyses will be needed.
Diatoms have the toolkit to metabolise amino acids and Dissolved Organic forms of Nitrogen (DON) and Berman and Bronk (2003) report evidence, even if occasional, of DON uptake by diatoms. Considering that DON concentration in the upper ocean layer is of several micromols/liter, of which 15-20% are urea or amino acids (Berman and Bronk, 2003), therefore from one to two order of magnitude larger than Dissolved Inorganic Nitrogen (DIN) in oligotrophic environments, it is possible to hypothesise that the diNPF transporters predicted to be localised to the plasma membrane, can have a very broad substrate specificity, such as amino acids. In fact, amino acids could provide alternative sources of energy in light-limiting conditions and alternative sources of N in nutrient-limiting conditions.

The NO$_3^-$ binding site in the plant AtNPF6.3 crystal structure includes a protonated histidine, His356 which forms an electrostatic interaction with the NO$_3^-$ molecule. Functional studies have shown that the point mutation of His356 in AtNPF6.3 resulted in the loss of function (Parker and Newstead, 2014). Interestingly, when Tyr370 of ZmNPF6.4 was mutated to His, the protein switched its preference to NO$_3^-$ over chloride (Wen et al., 2017). When His362 was mutated to Tyr in ZmNPF6.6, the protein could not transport NO$_3^-$ (Wen et al., 2017). These three functional studies indicate that histidine is essential for NO$_3^-$ transport. However, His356 is not conserved among AtNPF6.3 orthologues that harbour either a tyrosine or another hydrophobic residue at the corresponding position. diNPFs have a conserved tyrosine in the corresponding position (Fig. 2.8 and Table 2.6). Another tyrosine (or phenylalanine) from TMH1 is located in close proximity and may contribute to creating a hydrophobic pocket for the substrate (Fig. 2.8 and Table 2.6). The fact that all diNPFs lack this residue points to a different substrate than NO$_3^-$ (maybe peptides or chloride) or a different mode of substrate recognition. Functional studies are necessary to reveal the true role of NPFs in diatoms.

Yet, the short range of NO$_3^-$ conditions in the ocean could also be compatible with a dual affinity capacity of the diNPFs, as in the case of AtNPF6.3, OsNPF6.5, MtNPF1.3 and ZmNPF6.6 (Bagchi et al., 2012; Liu and Tsay, 2003; Wen et al., 2017) or with a preserved capacity to transport NO$_3^-$ even at low concentrations, a property that has been reported for some NPFs in plants (Valkov et al., 2017; Wang et al., 2020). In the case of AtNPF6.3 the switch between the two different modes of action in response to substrate availability, occurs through phosphorylation at threonine residue in position 101 (Liu and Tsay, 2003), that we found in some of the diNPF sequences (Table 2.6).
Interestingly, a significant percentage of sequences was predicted to be localised at the vacuole membrane, called tonoplast, and all these diNPFs sequences belong to the diNPFs Clade II. So, it is possible to hypothesise that diNPFs belonging to Clade II are not involved in uptake from the external environment but in reallocation of NO$_3^-$ between different cell compartments in response to specific N conditions. This would explain the conservation of transporters with low NO$_3^-$ affinity in diatoms despite its low concentrations in the ocean, because the NO$_3^-$ concentrations inside specific intracellular compartments can be much higher than in the external environment, reaching in some diatoms even up to 60 mM (Kamp et al., 2011). This would also be coherent with the lack of correlation between diNPFs gene expression and NO$_3^-$ availability in literature, in TARA Ocean database and in gene expression analyses performed on *P. tricornutum*.

In fact, diatoms can dominate phytoplankton communities and outcompete other eukaryotes under nutrient-limiting and nutrient-fluctuating environments, thanks to their ability to store nutrients, like NO$_3^-$, in large vacuoles during N repletion periods, and to make them available again when during N limited periods (Bromke et al., 2015). This storage capacity also plays an important role in relation to light: NO$_3^-$ can be assimilated into biomass and growth when light is present, and can be stored in the vacuole of diatoms to be exploited under darkness and anoxic conditions to gain energy, through Dissimilatory Nitrate Reduction to Ammonium (DNRA) which reduces intracellular NO$_3^-$ to NH$_4^+$ (Kamp et al., 2011, 2013).

Deciphering the complexity of the regulatory networks that control N uptake and metabolism will help understanding adaptation of diatoms to N availability in fluctuating intra- and extracellular environments and it will also provide new insights into the ecological success of these microalgae.

This Chapter laid the foundation for further functional analyses on the model diatom *P. tricornutum*, thanks to technologies enabling the generation and characterisation of overexpressing strains and knock-out mutants (Kroth et al., 2018). So, in the next Chapters, generation and characterisation of PtNPFs mutant strains will be described, in order to strengthen previous working hypothesis and to better understand PtNPFs role.

103
Chapter 3: Proteolistic protocol optimisation for *Ptnpfs* knock-out mutants generation

Abstract

The CRISPR/Cas9 system applied through proteolistics is a DNA-free nuclear transformation method based on the introduction of ribonucleoprotein complexes into cells. This method has been recently set up for diatoms and has several advantages, in particular by avoiding the occurrence of the transgene random insertions in the host genome and by limiting the working time of the Cas9 endonuclease, so reducing off-target mutations.

Here, proteolistic protocol was applied and optimised, increasing the gRNAs:Cas9 ratio from 1:1 to 6:1 with a consequent enhancement of mutation frequency, and decreasing the amount of Cas9 nuclease needed while maintaining a high efficiency comparable to the original protocol, that made it more cost-effective. Moreover, an increased concentration of the selective drug allows to reduce false positives.

Through this method, single and double knock-out mutants of PtNPF genes of the model diatom *P. tricornutum* were generated and clones with biallelic mutation were selected following an easy screening pipeline, for further phenotypic characterisation.
3.1. Introduction

Genetic engineering is a promising approach to obtain valuable insights into the diatom metabolism and physiology and to improve metabolic features for biotechnological applications (Faktorová et al., 2020; Serif et al., 2018). Advances in genetic engineering have generated molecular tools which can be used to modify genomes, overexpressing, knocking down and knocking out one or more genes of interest in several algal species (Sharma et al., 2021a).

Recently, the search for new techniques for diatom genome editing allowed the development of modern systems, such as TALEN (Transcription Activator-Like Effector Nucleases) (Daboussi et al., 2014) and CRISPR (Clustered Regulatory Interspaced Short Palindromic Repeats) / Cas9 (CRISPR Associated protein 9) (Nymark et al., 2016). In particular, the CRISPR/Cas9 system originally provides a sequence-specific adaptive immunity in bacteria, and acts by integrating short virus sequences in the cell CRISPR locus, allowing bacteria to remember, recognise and attack viral infections (Rath et al., 2015) (see paragraph 1.2.4 “Genome editing in diatoms” in Chapter 1).

This system was interestingly exploited as a two components knock-out approach based on the occurrence of a Double-Strand Break (DSB) induced by the endonuclease Cas9, in correspondence of a specific genomic DNA target recognised by a specific single-strand guide RNA (gRNA). The gRNA is made up of two parts: crispr RNA (crRNA), a 18-20 bp sequence complementary to the target DNA, and a trans-activating crispr RNA (tracrRNA), which is a universal binding scaffold for the Cas9. Through the guidance of gRNA, Cas9 cuts the DNA region located 10-12 bp proximal to a NGG sequence necessary for DNA cleavage, called Protopspacer Adjacent Motif (PAM) (Doudna and Charpentier, 2014). Then, DSB damage can be repaired by the cell through two different mechanisms (Fig. 3.1): Non-Homologous End-Joining (NHEJ) or Homology Directed Repair (HDR). NHEJ can induce INsertions and/or DELetions (INDELs) or single nucleotide replacements, often resulting in a frameshift and in the generation of a premature termination codon (PTC), with the consequent inactivation of the gene (Nymark et al., 2016; Kroth et al., 2018). While HDR can be used to make specific changes to the target region by providing a designed repair template that becomes inserted in the damaged region (Rath et al., 2015) (Fig. 3.1).
Figure 3.1. The mechanisms of CRISPR/Cas9-mediated genome editing and DSBs repair. CRISPR/Cas9 can introduce DSBs in DNA. DSBs are repaired by either NHEJ or HDR. Insertions, deletions or other alterations of DNA will occur during this process to achieve gene modifications. Image created with BioRender.

The CRISPR/Cas9 system can induce the occurrence of unwanted off-target mutations (Kroth et al., 2018), since the Cas9 can sometime cut at sites with a certain degree of similarity to the target region, and create undesirable perturbations. This happens when using the election technique for diatom transformation, the plasmid delivery through classical biolistic method: exogenous DNA, encoding both Cas9 and gRNA, attached to nanoparticles, is introduced by high pressure in target cells where it stably integrates in the endogenous genome (Falciatore et al., 1999; Poulsen and Kröger, 2005b). This established techniques, useful in overexpression gene studies, shows several disadvantages when applied to CRISPR/Cas9 system: low transformation efficiencies (less than $10^{-6}$ transformants per µg of DNA), long-term Cas9 expression which can induce off-target cleavage, and random integration of plasmid DNA into the genome, which can disrupt genes or alter gene regulation and expression near the integration sites (Serif et al., 2018).

In 2015, an alternative method has been set up. This method, called bacterial conjugation, consists in the introduction of a cargo DNA from *Escherichia coli* to diatoms throughout a conjugation process (Karas et al., 2015), and shows a reduced occurrence of off-target
mutations (Russo et al., 2018b; Moosburner et al., 2020). Bacterial conjugation still requires the introduction of exogenous DNA in the cells in the form of an episome, which is not integrated in the diatom genome and can be easily removed by eliminating antibiotic selection from the medium (Russo et al., 2018b). Although bacterial conjugation is a more precise and clean way than biolistics to introduce Cas9 nuclease into cells, it still elicits Cas9 expression over several generations, causing potential toxicity or off-target effects (Serif et al., 2018).

The recently introduced proteolitic method (Serif et al., 2018) is a DNA-, bacterial- and antibiotic-free technology. It is based on the use of RiboNucleoProtein complexes (RNP), formed by the nuclease Cas9 and genomic target-specific gRNAs. The most important advantages of this protocol are consequences of the Cas9 presence only as protein, and not as coding sequence integrated in the host genome: i) the avoidance of long-term expression of the nuclease in the cell, with the consequent the reduction of the off-targets, and ii) unwanted perturbations due to random integration of the transgene in the genome.

An additional advantage of the technology is the selection system, based on the simultaneous targeting of the gene(s) of interest along with the endogenous Adenine Phosphoribosyl Transferase (APT) and Uridine-5’-Monophosphate Synthase (UMPS).

APT normally contributes to adenine recycling as part of the nucleotide salvage pathway, when mutated, is able to confer resistance to 2-Fluoradenine (2-FA), a compound which is toxic for wild type cells (Schaff, 1994), when adenine is supplied to the medium. While, UMPS catalyses the critical step of the de novo pyrimidine biosynthesis pathway, essential for the production of uracil and conserved in all species (Sakaguchi et al., 2011). Its knock-out increases 5-fluoroorotic acid (5-FOA) tolerance, if uracil is supplied to the medium (Sakaguchi et al., 2011; Serif et al., 2018). In this way, the knock-out of APT or UMPS genes allows an antibiotic-free selection.

Here, this method was applied to the diatom Phaeodactylum tricornutum, firstly using two gRNAs for PtAPT as selection gene and then adding two other gRNAs for each target gene. 2-FA concentration was increased up to double compared to the original protocol (Serif et al., 2018), thus drastically decreasing the appearance of false positive colonies. Moreover, the protocol steps of RNPs complexes assembly were modified being able to decrease of around 80% the amount of Cas9 protein required for each shot, while increasing the gRNAs:Cas9 ratio and maintaining a high transformation efficiency (Russo et al., 2022).
3.2. Materials and Methods

3.2.1. CRISPR-Cas9 design and constructs

Before designing and ordering the gRNAs, full-length *PtNPF1* and *PtNPF2* candidate genes were amplified and sequenced from *P. tricornutum* strain, to confirm their sequences did not differ from those reported in the reference genome ([https://protists.ensembl.org/Phaeodactylum_tricornutum/Info/Index?db=core](https://protists.ensembl.org/Phaeodactylum_tricornutum/Info/Index?db=core)).

gRNAs were selected using the CRISPOR tool ([http://crispor.tefor.net/](http://crispor.tefor.net/)) (Haeussler et al., 2016) based on the Moreno-Mateos score and the absence of predictable off-targets. Other online tools can be used to design gRNAs, such as CHOPCHOP ([https://chopchop.cbu.uib.no/](https://chopchop.cbu.uib.no/)) and phytoCRISP-Ex ([https://www.phytocrispex.biologie.ens.fr/CRISP-Ex/](https://www.phytocrispex.biologie.ens.fr/CRISP-Ex/)). In particular, these tools identify Cas9 target sequences, namely crRNAs, defined by the protospacer (20 bp) and the PAM sequence, which need to be subsequently linked to the universal tracrRNA to form complete gRNAs.

Some basic criteria were followed: i) crRNAs should preferably be located at the 5’ end of the gene, in correspondence of exons; ii) sequences containing single nucleotide polymorphisms (SNPs) between alleles should be avoided to prevent targeting a preferential allele over the other; iii) two target sequences for the same gene were chosen to be used simultaneously, in order to increase the probability to generate a frameshift as a result of an induced INDEL.

Considering features of *PtNPF* genes, their crRNAs were chosen based on the protein structural models, selecting sequence regions that corresponded to transmembrane domains involved in significant bonds. Corresponding crRNA sequences targeting *PtNPF1* (ID: Phatr3_J47148, Supplementary File S3.1) were gNPF1_a and gNPF1_b (Table 3.1). crRNA sequences targeting *PtNPF2* (ID: Phatr3_J47218, Supplementary File S3.1) were gNPF2_a and gNPF2_b (Table 3.1). In both cases, the gene region between the two gRNAs designed included key residues involved in the protein folding or substrate recognition or protein conformation stability.
In addition, \textit{PtAPT} (ID: Phatr3\_J6834, Supplementary File S3.1) was used as selectable markers, with gRNAs (\textit{gAPT}\_1 and \textit{gAPT}\_3) previously designed by Serif et al. (2018) and targeting regions within exons 1 and 2, respectively (Table 3.1).

\textbf{Table 3.1.} List of crRNAs designed for CRISPR/Cas9 proteolistic transformation, with the PAM sequence underlined.

<table>
<thead>
<tr>
<th>crRNA</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{gAPT}_1</td>
<td>5’-AAGCGTGGAATGCCCTTTGAAGGG-3’ (-)</td>
</tr>
<tr>
<td>\textit{gAPT}_3</td>
<td>5’-CCAGGGCAATTGTGGGACCCAGG-3’ (-)</td>
</tr>
<tr>
<td>\textit{gNPF1}_a</td>
<td>5’-AAAGGGAGAGGTGTAGGCAACGG-3’ (-)</td>
</tr>
<tr>
<td>\textit{gNPF1}_b</td>
<td>5’-GGCATGCTGTCCTTGGGAGTCGG-3’ (+)</td>
</tr>
<tr>
<td>\textit{gNPF2}_a</td>
<td>5’-TGTTGTGCTAGGAACAGGAGG-3’ (+)</td>
</tr>
<tr>
<td>\textit{gNPF2}_b</td>
<td>5’-TACGAGTGCAATCGATGTGACCG-3’ (-)</td>
</tr>
</tbody>
</table>

\textbf{3.2.2. crRNA::tracrRNA duplexes forming gRNAs}

Lyophilised CRISPR-Cas9 RNAs (crRNAs) and CRISPR-Cas9 trans-activating RNA (tracrRNA) were purchased from IDT and resuspended in Nuclease-Free Duplex Buffer (IDT - Integrated DNA Technologies) mixing up and down to a final concentration of 100 µM. Each crRNA::tracrRNA solution was then prepared in order to contain 30 µM duplex, mixing together 3 µl of 100 µM CRISPR crRNA, 3 µl of 100 µM CRISPR tracrRNA, 4 µl Nuclease-Free Duplex Buffer (IDT - Integrated DNA Technologies). The mix was incubated for 5 minutes at 95 °C in a thermocycler and immediately removed from the block, cooling down at room temperature on the bench (Fig. 3.2) and then stored at -20°C.
3.2.3. Diatom culture and plating

The axenic CCMP632 strain of *P. tricornutum* Bohlin was obtained from the Provasoli-Guillard National Center for Culture of Marine Phytoplankton. It corresponds to the ecotype Pt1 (De Martino *et al.*, 2007; Rastogi *et al.*, 2019). *P. tricornutum* was grown in autoclaved 0.22 μm filtered F/2 medium without silica (Guillard, 1975) at 18 °C under 12 h:12 h day:night cycles, 90 μmol photons m$^{-2}$ s$^{-1}$ white light during the day, in order to have 1.5 x 10$^8$ total cells per shot, collected during the exponential phase (around 1.0-2.0 x 10$^6$ cells/ml), on the day before the shots. When the cells have reached the desired
concentration, the culture volume required to get $1.5 \times 10^8$ cells was collected by centrifugation at 1500 g at 18 °C for 10 minutes, the supernatant was discarded, the pellet was resuspended with the bit of supernatant left and the cells spread on a plate forming a circle of around 4 cm in diameter. Plates with F/2 medium containing deionized sterile water, 20 g/l Sea salts, F/2 solution 1X without silicate (Guillard, 1975) and 10 g/l agar bacteriological grade.

### 3.2.4. Assembly of RNP complexes

The needed amount of HiFi Cas9 Nuclease V3 protein was diluted from a first stock solution at 10 g/l (i.e. 65.5 µM), to a second stock solution at 2.5 g/l (i.e. 15.6 µM) using nuclease free H$_2$O. This step facilitated downstream work because the solution at 10 g/l is viscous and hard to manipulate [Molecular Weight of Cas9 = 1.6 x 10$^5$ g/mol].

Proteolistic experiments to knock-out PtAPT using only its two gRNA, and to generate single PtNPF knock-out mutants using two gRNA for the target gene and two for the selection gene, were performed strictly following Serif et al. (2018). According to this protocol each gRNA was assembled with Cas9 protein separately, and then put together just before the transformation, in order to have four gRNAs together with 8 µg of Cas9 in a final volume of 8 µl per shot (Table 3.2).

Then, the protocol was optimised to generate double PtNPFs knock-out mutants, assembling the six gRNAs with total 1.5 µg Cas9 in only one step. In particular, two gRNAs for each of two target genes (PtNPF1 and PtNPF2) and two gRNAs for the endogenous selection gene (PtAPT) were included (Fig. 3.2 and Table 3.2). Moreover, for each RNP complex to assemble, all the components were mixed together in order to have 1.5 µg of Cas9 in a final volume of 8 µl per shot.

**Table 3.2.** Mix of components for each RNP complexes to assemble, specific for the transformation experiments performed. For each experiment, the total Cas9 amount has been calculated.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Component</th>
<th>Volume (µL)</th>
<th>Total Cas9 amount (µg)</th>
<th>Molar ratio gRNAs:Cas9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ptapt knock-out</td>
<td>APT_1 gRNA</td>
<td>1.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>APT_2 gRNA</td>
<td>1.6</td>
<td>8</td>
<td>1:1</td>
</tr>
<tr>
<td></td>
<td>Alt-R Cas9 enzyme (at 15.6 µM)</td>
<td>3.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The first two experiments were performed with 8 µg Cas9 per shot and a gRNAs:Cas9 molar ratio of 48 pmol gRNA : 49 pmol Cas9 (1:1) such as the original protocol (Serif et al., 2018). The last experiment was optimised to exploit only 1.5 µg Cas9 per shot and a gRNAs:Cas9 molar ratio of 54 pmol gRNA : 9 pmol Cas9 (6:1).

For the negative control, with only Cas9, the same amount of enzyme of the other shots was used: 0.6 µl diluted Cas9, 7.4 µl PBS (1X).

The reaction was incubated at room temperature for 20 min, then RNP complexes were stored on ice.

### 3.2.5. Cas9 and gRNAs in vitro functional validation

Genomic fragments encompassing the gRNA region of selection and target genes were amplified with the primers for *PtAPT*, *PtNPF1* and *PtNPF2* describer in Table 3.3. Fragments were then purified using the QIAquick® PCR Purification Kit (Qiagen) and following manufacturer instructions.
Table 3.3. List of primers used to amplify target genes fragments, in order to test designed gRNAs.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PtAPT_fl_for</td>
<td>5'-ATGACGACAAACCAACGGAAG-3'</td>
</tr>
<tr>
<td>PtAPT_fl2_rev</td>
<td>5'-ATTTTGACCAAATAACTACCCGGAAG-3'</td>
</tr>
<tr>
<td>PtNPF1_fl_for</td>
<td>5'-ATGACGACTCCCACTACGGC-3'</td>
</tr>
<tr>
<td>PtNPF1_fl_rev</td>
<td>5'-CTAGGAGGACGTTGCTCGCA-3'</td>
</tr>
<tr>
<td>PtNPF2_fl2_for</td>
<td>5'-CATGCCCAGGCAACGGAACGACC-3'</td>
</tr>
<tr>
<td>PtNPF2_fl_rev</td>
<td>5'-TCAAATCTATTCTTCGCTC-3'</td>
</tr>
</tbody>
</table>

The Cas9-RNP complexes were mixed with the purified DNA fragment, with 10:1 molar ratio in 10 µl final volume with PBS (1X), considering to bring DNA final concentration to 50 nM. The reaction was incubated at 37 °C for 60 min, add 20 µg proteinase K and incubate at 56 °C for 10 min. The digestion of the PCR product was analysed by agarose gel electrophoresis to verify the presence of fragments deriving from the Cas9 cut.

3.2.6. Preparation of gold nanoparticles and loading of RNP complexes onto them

Gold nanoparticles are the microcarriers on which RNPs can be loaded. 60 mg of dry gold nanoparticles (0.6 µm diameter) were placed in 1 ml of 100% ethanol in a microfuge tube, vortex on high for 1 minute and centrifuge at max speed, room temperature for 1 minute. The washing steps were repeated twice, and particles were pelleted again at max speed, room temperature for 1 minute. Two washes were done in 1 ml sterile distilled water. Nanoparticles were resuspended in 1 ml sterile water and aliquot 50 µL into sterile microfuge tubes, vortexing continuously while pipetting. Each 50 µL aliquot contains 3 µg gold particles, sufficient for 5 proteolistic shots and can be stored at -20 °C for months.

The transformation day, gold nanoparticles aliquots were thawed at room temperature. They were centrifuged at maximum speed, room temperature for 1 minute, the supernatant removed, and the pellet resuspended with 50 µL PBS (1X). After two washing steps, the pellet was resuspended with 50 µL PBS (1X) by flickering and vortexing. 10 µL of the homogeneous gold nanoparticles solution were transferred into each RNP mixture, so allowing RNPs loading on microcarriers. Then, the solution was vortexed and spread at the centre of the PDS-1000/He macrocarriers (BioRad), previously washed with 100% ethanol and placed into PDS-1000/He macrocarrier holders (BioRad) (Fig. 3.2).
3.2.7. Transformation and replating

Proteolistic protocol occurs through a biolistic shot, performed using a Gene-gun PDS-1000/He System and according to (Falciatore et al., 1999).

In detail, the main valve of the helium cylinder was open until gas pressure registered on the first gauge of the regulator. Then, the regulator was turned adjusting screw clockwise and adjusting the pressure to 1700 psi (200 psi above the burst pressure of the selected rupture disk). The 1550 psi rupture disk was put inside the assembly and screwed onto the acceleration tube. The macrocarrier holder, containing the microcarrier and the dried RNP complex-coated nanoparticles, was assembled with a sterile stopping screen and installed with facing down. The whole assembly was placed on the second shelf slot in the chamber, while the Petri dish containing cells on the dish holder was placed on the third shelf slot, and the bombardment chamber door was closed (Falciatore et al., 1999).

With the power on, the vacuum pump started and, when the vacuum gauge showed 23 in Hg, the switch was turned to Hold, and the fire switch was pressed continuously until the rupture disk burst and the helium pressure gauge dropped to zero.

Then the central switch was turned to Vent to release the vacuum, the bombardment chamber door was open and the plate removed, unloading the macrocarrier and stopping screen from the microcarrier launch assembly, and unloading the spent rupture disk.

Immediately after the shots, all the Petri dishes were placed in the growth chamber, cells facing down, 18 °C under 12 h:12 h day:night cycles, 90 µmol photons m$^{-2}$ s$^{-1}$ white light during the day (Falciatore et al., 1999).

After 48h from the shots, cells were resuspended with 1 ml sterile F/2 medium without silicate and spread equally between two selective plates (Fig. 3.2). Selective plates were done with F/2 medium without silicate, containing 20 µM 2-FA as drug and 5 µg/ml adenine as supplement. Plates were put in the growth chamber, cells facing down, at 18 °C, 12 h:12 h day:night cycles, 100 µmol photons m$^{-2}$ s$^{-1}$. Resistant colonies appeared after 2-3 weeks.

Only in case of the consecutive proteolistic experiment, with both PtAPT and PtUMPS selection genes exploited, 600 µg/ml 5-FOA as drug and 5 µg/ml uracil as supplement were added to selective plates.
All of the appropriate controls were included, comprising a negative control by transforming cells with Cas9 alone and a positive control with RNPs targeting the endogenous selection gene alone (PtAPT).

3.2.8. PCR analysis on transformed knock-out cells and Sanger sequencing

PCR was performed on genomic DNA from transformant and wild-type cells, picked from selective and control plates. Cells were treated with a Lysis Buffer consisting of 1% (v/v) Triton X, 20 mM Tris-HCl (pH 8) and 2 mM EDTA (pH 8) in Milli-Q water, vortexed, put on ice for 15 minutes and then at 85°C for 10 minutes, as described in Daboussi et al. (2014). The diluted solution was then used as template for PCR with the following thermal profile: 95°C for 1 min, (95°C for 20 sec, 58°C for 20 sec, 72°C for 2 min) ×35 cycles, 72°C for 2 min. Primers used are described in Table 3.4.

Table 3.4. List of primers used for the screening, amplifying fragments on the selective gene PtAPT and on the target genes PtNPF1 and PtNPF2.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Fragment length</th>
</tr>
</thead>
<tbody>
<tr>
<td>PtAPT_fl_for</td>
<td>5’-ATGACGACAACCAACCGGAAG-3’</td>
<td>771 bp</td>
</tr>
<tr>
<td>PtAPT_fl_rev</td>
<td>5’-TCAGTGTTCCTCGCCGTCATC-3’</td>
<td></td>
</tr>
<tr>
<td>PtNPF1_ko_for</td>
<td>5’-AACTGCTCGAGCGATTTTCC-3’</td>
<td>675 bp</td>
</tr>
<tr>
<td>PtNPF1_ko_rev</td>
<td>5’-AATCCGTGTGGGGAGGGGT-3’</td>
<td></td>
</tr>
<tr>
<td>PtNPF1_fl_for</td>
<td>5’-ATGACGACTCCCAGTGAGA-3’</td>
<td>1839 bp</td>
</tr>
<tr>
<td>PtNPF1_exp_rev</td>
<td>5’-ATCCGACTCCCAGTGAGA-3’</td>
<td></td>
</tr>
<tr>
<td>PtNPF2_ko_for</td>
<td>5’-ACGTTACCAACAAATCTGATG-3’</td>
<td>645 bp</td>
</tr>
<tr>
<td>PtNPF2_ko_rev</td>
<td>5’-GTGTGTGCTGAGATGGGATCC-3’</td>
<td></td>
</tr>
<tr>
<td>PtNPF2_exp_for</td>
<td>5’-TTACGTGATTTGCTTGTCCA-3’</td>
<td>1573 bp</td>
</tr>
<tr>
<td>PtNPF2_fl_rev</td>
<td>5’-TCAAATCATTTCTCGCTC-3’</td>
<td></td>
</tr>
</tbody>
</table>

The Q5™ High Fidelity DNA polymerase (New England BioLabs, USA) was used in order to characterise INDELs length. When a single product was obtained from PCR, it was purified using QIAquick® PCR Purification Kit (Qiagen) and directly sequenced; while when two products were obtained, amplicons were subcloned using TOPO TA Cloning™ (Invitrogen). Sanger sequencing was performed on an Applied Biosystems (Life Technologies) 3730 Analyzer 48 capillaries using the same primers used to generate
the PCR product or M13_for 5'-GTAAAACGACGGCCAG-3' and M13_rev 5'-CAGGAAACAGCTATGAC-3' present at both sides of TOPO TA vector cloning site.

Single Nucleotide Polymorphisms (SNPs) flanking mutations in \textit{PtNPFs} gene were used to distinguish alleles in \textit{P. tricornutum} (Sharma et al., 2018). After sequencing, mutated sequences were \textit{in silico} translated with the online tool Expasy Translate (https://web.expasy.org/translate/) and aligned with wild-type amino acid sequence to identify frameshifts and stop codon insertions. Then, structural models of truncated proteins were reconstructed with the online tool SWISS-MODEL (https://swissmodel.expasy.org/) to visualise them.

Finally, cell lines with biallelic mutations both resulting in truncated proteins were selected for further experiments and phenotypic characterisation.

\section*{3.3. Results}

\subsection*{3.3.1. Drug selection set up}

\textit{P. tricornutum APT} was chosen as main selectable marker for proteolistic transformations (Serif et al., 2018). \textit{APT} encodes a purine salvage enzyme that catalyses the conversion of adenine and phosphoribosyl pyrophosphate into AMP, so normally contributing to adenine recycling as part of the nucleotide salvage pathway (Gaillard, 1998). As previously said, the inactivation of this gene confers 2-FA resistance in several organisms (Lamichhane-Khadka et al., 2021), including \textit{P. tricornutum} (Serif et al., 2018).

The 2-FA drug efficacy was tested on wild-type \textit{P. tricornutum} strains Pt1 and a \textit{Ptapt} knock-out mutant previously generated by Serif et al. (2018). Increasing 2-FA concentrations were tested both on plates and in liquid cultures: no drug, 10, 20, 30 and 50 µM (Fig. 3.3). While the original protocol used 10 µM for selection, at this concentration some wild-type residual cells were observed both on plates and liquid cultures, so selective 2-FA concentration was increased up to 20 µM, to further reduce the risk of false positive colonies growth after transformations. To note that adenine have to be added to \textit{Ptapt} knock-out mutant cultures to allow their survival and growth, as consequence of the inactivation of the adenine recycling cell system.
Figure 3.3. *P. tricornutum* Pt1 wild-type and *Ptapt* exposed to different 2-FA concentrations. Cells plated on solid F/2 medium without silica, with adenine (ADE) and with A) 50 µM, B) 30 µM, C) 20 µM, D) 10 µM and E) no 2-FA; F) cells growing in liquid F/2 medium without silica, with adenine and with 20 µM, 10 µM and no 2-FA.

### 3.3.2. gRNAs design and *in vitro* Cas9 assays

The Cas9 endonuclease has the capability to cut a DNA sequence by generating a DSB. The sequence to be cut is recognised by a complementary RNA strand called crRNA, that together with a universal tracrRNA forms a gRNA. Before the proteolistic transformation, it is necessary to ensure that the RNP complex, formed by the Cas9 and the gRNA, is able to recognise and cut the target sequence. So, the enzymatic activity of Cas9 and the recognition of the target genes by the designed crRNAs were tested through an *in vitro* assay, that involves the digestion of the target sequence previously amplified by PCR reaction (see before Materials and Methods paragraph 3.2.5 “Cas9 and gRNAs *in vitro* functional validation”).

In particular, two gRNA were designed for each gene, in order to improve the mutation frequency, which was found to be higher than that induced by a single gRNA (Hao et al., 2021).

As *PtAPT* was chosen as main selectable marker for proteolistic transformations, the *in vitro* Cas9 assay was first set up using two *PtAPT* gRNAs previously designed by Serif.
et al. (2018). The 560 bp *PtAPT* target sequence, including both selected crRNAs, was amplified by PCR and then combined with RNP complexes, for gAPT_1 and gAPT_3 gRNAs, respectively.

The expected Cas9 cleavage products migrated onto agarose gel are shown in Figure 3.4, confirming the enzymatic activity of Cas9 and the target sequence recognition by crRNAs. In particular, gAPT_1 cuts *PtAPT* amplicon into two fragments of 455 bp and 105 bp, while gAPT_3 cuts it into two fragments of 225 bp and 335 bp.

![Figure 3.4. *in vitro* Cas9 assay with *PtAPT* gRNAs: gAPT_1 cuts *PtAPT* amplicon into two fragments of 455 bp and 105 bp, while gAPT_3 cuts it into two fragments of 225 bp and 335 bp. M represents the 100 bp marker.](image)

Then, two gRNAs directed against *PtNPF1* (gNPF1_a and gNPF1_b) were designed to target a gene region which corresponds to transmembrane helices in folded proteins. In particular, the two gRNAs designed cover the 715th-1124th bp region of *PtNPF1* gene sequence, which corresponds to the 238th-374th aa region of PtNPF1 protein sequence (Fig. 3.5A). This region includes four transmembrane helices (from TMH2 to TMH6) (Fig. 3.5B), in which there is one of the key residues K310, involved in the salt-bridge that stabilise outward open conformation of PtNPF1 protein (see paragraph 2.3.7. “Structural modelling of diNPFs” in Chapter 2).
In vitro Cas9 assay was performed on *PtNPFI* amplified from genomic DNA, to test gRNAs efficiency. An amplified 2328 bp *PtNPFI* fragment was combined respectively with gNPFI\_a, obtaining two fragments of 1613 bp and 715 bp, and with gNPFI\_b, obtaining two fragments of 1204 bp and 1124 bp (Fig. 3.6A). Moreover, to better simulate next proteolistic experiments involving both selection and target genes, RNP complexes made combining both *PtAPT* and *PtNPFI* gRNAs (so containing gAPT\_1, gAPT\_3, gNPFI\_a and gNPFI\_b) were used to perform in vitro Cas9 assay on *PtAPT* and *PtNPFI* fragments respectively. Both fragments were cut by their corresponding gRNAs, showing four amplicons on agarose gel (Fig. 3.6B).
Then, gRNAs directed against *PtNPF2* (gNPF2_a and gNPF2_b) were designed and *in vitro* Cas9 assay was performed. Also in this case, gRNAs were designed in order to target a region corresponding to transmembrane helices in folded protein: the two gRNAs cover the 500\textsuperscript{th}-829\textsuperscript{th} bp region of *PtNPF2* gene sequence, which corresponds to the 166\textsuperscript{th}-276\textsuperscript{th} aa region of *PtNPF2* protein sequence (Fig. 3.7A). This region includes two transmembrane helices (TMH4 and TMH5) (Fig. 3.7B), in which there is one of the key residues K169, involved in the salt-bridge that stabilise outward open conformation of *PtNPF2* protein, similarly to gRNAs previously designed for *PtNPF1* (see paragraph 2.3.7. “Structural modelling of diNPFs” in Chapter 2).
**Figure 3.7.** gRNAs designed and correspondence to TMH4 and TMH5 in the folded protein. A) yellow bars indicate the gRNAs located on the *PtNPF2* gene sequence with bold underlined PAM sequence, B) predicted *PtNPF2* protein structure, with residues included between the two designed gRNAs in red.

*In vitro* Cas9 assay was performed on *PtNPF2* amplified from genomic DNA, to test the efficiency of the designed gRNAs. An amplified 1861 bp *PtNPF2* fragment, corresponding to complete gene sequence, was combined respectively with gNPF2_a, obtaining two fragments of 1456 bp and 405 bp, and gNPF2_b, obtaining two fragments of 1127 bp and 734 bp (Fig. 3.8).

**Figure 3.8.** *In vitro* Cas9 assay with *PtNPF2* gRNAs: gNPF2_a cuts *PtNPF2* into two fragments of 1456 bp and 405 bp, while gNPF2_b cuts *PtNPF2* into two fragments of 1127 bp and 734 bp. M represents the 1Kb marker.
3.3.3. Pipeline for proteolistic mutants generation and screening

After each transformation, *P. tricornutum* colonies grown on selective plates were transferred on a new one and in liquid medium. Then, colonies screening was performed following different steps: i) genomic DNA extraction of clones positive to selection, PCR screening and visualisation of amplicons at different height on the agarose gel, ii) eventual cloning of single amplicons if a multiple-amplicon pattern needs to be separated, iii) Sanger sequencing, iv) visualisation of chromatograms and alignment with wild-type reference gene, v) SNPs study when appropriate, vi) prediction of resulted mutated protein (Fig. 3.9).

First of all, a PCR has to be performed on genomic DNA using primers which amplify a fragment including both gRNA. Through the first PCR screening, large INDELs can be immediately detected from agarose gel visualisation (Figs. 3.9A-C and Table 3.3 for primers), in particular amplicons lower than wild-type represent deletions while higher amplicons indicate insertions. Through this analysis, transformed strains can show a different amplicon pattern with one or more amplicons visible on the agarose gel, indicating different types of mutation (Fig. 3.10).

Figure 3.9. Pipeline for proteolistic mutants screening. A) Collection, lysis and gDNA extraction of clones positive to selection, B) PCR screening and C) visualisation of amplicons at different height on agarose gel, D) eventual cloning of single amplicons, E) Sanger sequencing, F) visualisation of chromatograms and G) alignment with wild-type reference gene, H) eventual SNPs study, I) prediction of resulted mutated protein. All these steps are described in detail in the text below.
This first PCR screening help to fast identify interesting mutations, which subsequently need to be better characterised through direct Sanger sequencing. In some cases, it could be also useful to clone amplicons before sequencing, in order to separate them when a multiple-amplicons pattern is present (Figs. 3.9D-F).

Resulted chromatograms of wild-type and mutated strains can be compared, to clarify the mutation width (Fig. 3.9G). So that the subsequent step of knock-out screening pipeline would use obtained chromatograms to study mutation flanking SNPs, in order to understand if a mutation is mono- or biallelic (Fig. 3.9H).

Considering the different amplicon patterns visible as the result of the first screening PCR, three main scenarios can be observed: strains showing i) only one amplicon of the same or different weight of wild-type, ii) two amplicons, of which both differ from wild-type or one can present the same wild-type weight, and iii) three amplicons (Fig. 3.10).

In detail, strains showing a single amplicon lower or higher than wild-type (e.g. Fig. 3.10 clone 1) indicate a biallelic mutation, which could be homozygous or heterozygous in case of a second larger INDEL too large to be amplified by PCR.

Further characterisation of these mutation can be achieved by analysing mutation-flanking SNPs (Fig. 3.9H). They are substitutions of a single nucleotide at a specific position in the genome that commonly occur in wild-type P. tricornutum (more than 460 000 SNPs - Rastogi et al., 2020): if SNPs are observed in both wild-type sequence and mutant, the mutation is homozygous; while the absence of SNPs in the mutant could indicate a large deletion on one of the alleles which was not amplified through diagnostic PCR, or a phenomenon called Loss Of Heterozygosity (LOH). LOH is a cross chromosomal event that results in the loss of the entire wild-type gene and the surrounding chromosomal region, and can arise from the repair mechanisms used by organisms in response to DNA double-strand breaks (DSBs) (Feri et al., 2016; Joseph et al., 2014).

Also strains presenting an amplicon of the same weight of wild-type could be mutated, with small INDELs of few nucleotides which could not be visible from a 1% agarose gel run, and that need to be sequenced to be detected.

In case of strains showing two amplicons from the agarose gel, they could be monoallelic mutations if one of the amplicons corresponded to wild-type, or biallelic mutations if both alleles differed from wild-type (e.g. Fig. 3.10 clone 3).
It is also possible to observe three amplicons on the agarose gel (e.g. Fig. 3.10 clone 6): in this case the colony could be the result of a mixed population of strains differently mutated, and it needs one or more sub-cloning step to isolate single clones. But if the three-amplicon mutation pattern occurs also in subclones the reason can be different.

A first explanation can be the presence of pseudogenes, which are other copies of a specific gene that have lost their function (often because of rearrangements) and cannot make RNA. CRISPR/Cas9 can target all the gene copies, as long as the gRNA-recognised sequence is there, but different repair mechanisms occurred after DSB on different gene copies can cause a multiple-amplicons pattern during the PCR screening. This is not the case of PtNPF genes, that are present as single copy genes in P. tricorntum genome. Another explanation can be given by the formation of a heteroduplex caused during late stages of PCR reactions, when high levels of products and other reactants, such as primers and polymerase, are depleted. Here, two strands differently mutated can anneal forming a bubble of single-strand DNA in the middle, and this bubble lead the heteroduplex to move in the gel at an unusual size (Delwart et al., 1995).

Figure 3.10. Example of proteolistic mutants screening by PCR. Different amplicon patterns can be observed on the resulting agarose gel: they indicate that different hypothetical mutations occurred, which need to be subsequently tested through sequencing. M represents the marker.

Once the mutation width is identified, the mutated sequence can be in silico translated and compared with the wild-type one: this step allows to highlight frameshifts which
resulted in the amino acid insertion, deletion or change, including stop-codon insertion which cause the translation block (Fig. 3.9I). Most interesting strains, selected for subsequent phenotypic characterisation, may have mutations which resulted in truncated proteins.

3.3.4. Ptapt knock-out mutants generation and screening

To reproduce and optimise the proteolistic protocol on P. tricornutum (Serif et al., 2018), a first set up was done using only the two gRNAs targeting the selection gene PtAPT (Supplementary File S3.1). Adapting Serif et al. (2018) protocol as described before in Materials and Methods paragraph, proteolistic shots gave more than total 100 colonies positive to 2-FA selection (Figs. 3.11A). Five colonies were screened by PCR in order to detect INDELs, observing three of them with large insertions or mutations clearly visible on the agarose gel (Fig. 3.11B and Supplementary Table S3.1). All the amplicons were sequenced to define the INDEL magnitude of mutated clones, but also to check eventual small mutations in clones showing a single amplicon of the same wild-type weight. The clone 3 was found to be equal to wild-type with no mutations while, in Figure 3.11C, clone 4 mutation is reported: a 541 bp deletion occurs between the two gRNAs designed.
Figure 3.11. Selection and identification of Ptapt proteolistic mutants. A) Positive C+ and negative C- control plates with non-transformed wild-type cells growing on F/2 medium without and with selection respectively, and two plates with transformed cells growing on selection. B) agarose gel showing the PCR screening on transformed cells positive to selection: M represents the 100 bp marker and C- the blank. C) Ptapt clone 4 chromatogram showing deletion: Cas9 cut genomic DNA in correspondence of the two designed gRNAs (underlined, with PAM sequence in the black squares), the green and pink shadows represent the upstream and downstream regions where wild-type and mutant strains correctly align.

*PtAPT* wild-type gene is 771 bp long. The 230 bp deletion in Ptapt clone 4 (Supplementary Table S3.1 and Supplementary File S3.2) causes a frameshift from 108th nucleotide, corresponding to the 36th amino acid, with the consequent insertion of a stop codon which blocks protein translation. The model of the resulting truncated protein is represented in Figure 3.12.
Figure 3.12. *Ptapt* clone 4 mutation resulting in a truncated protein. A) Translated mutated sequence aligned with wild-type reference amino acid sequence, showing the translation halt; B) predicted protein structures of wild-type and *Ptapt* clone 4 APT proteins.

An interesting case is represented in Figure 3.11B, clone 1, which shows three amplicons. One of the proteolistic advantages is that Cas9 protein is not maintained in cells after mitotic division, so reducing the need for subsequent subcloning steps to isolate single clones from a different cell population. Anyway, considering *P. tricornutum* is a biallelic microorganism and no multiple *PtAPT* copies were retrieved in *P. tricornutum* genome, the cells were subcloned on medium containing 2-FA and adenine to try to isolate individual clones. After subcloning, all the strains isolated presented the same original three-amplicons mutation pattern (Fig. 3.13), and while the higher and lower amplicons were easy to sequence, revealing a wild-type allele and a 541 bp deleted allele, the middle amplicon presented sequencing errors or not clearly define peaks on the chromatograms. The hypothesis is that clone 1 is not a mixed population of different clones, but that the three-amplicons pattern could be due to the formation of a heteroduplex, which migrates differently on the agarose gel (see previous paragraph 3.3.3 “Pipeline for proteolistic mutants generation and screening”).
Figure 3.13. *Ptapt* clone 1 subclones further screened by PCR, still highlight three amplicons on agarose gel. M represents the 100 bp marker and C- the blank.

Through this first proteolistic transformations and colonies screening, the protocol was set up and the mutations of the *PtAPT* selection gene were verified. These experiments laid the foundations for subsequent transformations, that used guides for target genes in addition to those for the selection gene.

3.3.5. *Ptapt*-*PtNpf1* knock-out mutants generation and screening

To study the role of NPFs in *P. tricornutum* physiology and viability, CRISPR/Cas9 genome editing through proteolistic method was used to obtain knock-out strains. A proteolistic transformation was performed, using the two gRNAs targeting *PtAPT* and the two gRNAs targeting *PtNPF1* (Supplementary File S3.1).

After transformation, almost more than total 200 colonies positive to selection were obtained from each shot, and total 36 clones for *Ptapt*-*PtNpf1* knock-out were selected for further screening: 55.5 % of these clones presents INDELs on the target gene (Figs. 3.14A-B and Supplementary Table S3.1).
Figure 3.14. Selection and first PCR screening of *Ptnpf1* proteolistic mutants. A) Positive C+ and negative C- control plates with not transformed wild-type cells growing on F/2 medium without and with selection respectively, and four plates with transformed cells growing on selection. B) agarose gels showing the results of the PCR screening on *PtAPT* and *PtNPF1* genes on wild-type and transformed cells positive to selection, M represents the 100 bp marker (above and the 1Kb marker (below), while C- represents the blank.

Amplicons of most interesting strains were sequenced to define the INDEL magnitude. Clones showing only one or two amplicons of a different weight than wild-type were chosen for Sanger sequencing. Moreover, for single amplicon clones, mutation flanking SNPs were studied, in order to define whether the biallelic mutations were homo- or heterozygous. As example, in Figure 3.15B, mutation flanking SNPs were analysed in
wild-type and Ptnpf1 clone 2.6: this clone shows a single amplicon from agarose gel (Fig. 3.14B), lower than wild-type, which corresponds to a 404 bp deletion. In wild-type, SNPs flanking this genomic DNA region, are present on the chromatogram as double peaks, indicating different nucleotides on different alleles, while in Ptnpf1 clone 2.6, chromatogram shows only one peak in correspondence of the wild-type SNPs, indicating that mutation is present only on one allele (Fig. 3.15B). This suggested that Ptnpf1 clone 2.6 presents a PtNPF1 biallelic heterozygous mutation, of which one deletion of known length while the other not yet defined. The latter mutation could have a much greater width and could involve other regions of the genome flanking PtNPF1, which might be important in other processes. As alternative, this strain could be the result of a LOH event.

Figure 3.15. A) PtNPF1 gene on P. tricornutum wild-type genome showing SNPs indicated by red arrows. B) Comparison between wild-type chromatogram showing SNPs and Ptnpf1 mutant chromatogram which does not present SNPs up- and downstream of the mutation.

In addition to clone 2.6, four other Ptnpf1 clones presenting a single amplicon at a height different from wild-type on the agarose gel, were sequenced and the mutation-flanking SNPs analysed: all of them show heterozygous biallelic mutations and none of these was chosen for further analyses and phenotypic characterisation.

On the other hand, Ptnpf1 clones presenting two amplicons, and so indicating a heterozygous biallelic mutation were easier to identity in term of width. In particular, Ptnpf1 clones 2.8 and 2.9 show two amplicons from agarose gel (Fig. 3.14B), which were sequenced: Ptnpf1 clone 2.8 highlights an insertion of 214 bp on one allele and a deletion
of 403 bp on the other one, while *Ptnpf1* clone 2.9 presents two deletions of 403 and 650 bp, respectively (Figs. 3.16B-C and Supplementary File S3.3).

**Figure 3.16.** *P. tricornutum* wild-type and *Ptnpf1* mutants schemes and chromatograms showing Cas9 different cuts on the two *Ptnpf1* alleles. A) *PtNPF1* wild-type sequence showing gRNAs and primer couples used for sequencing. B) *Ptnpf1* 2.8 mutation scheme and comparison between wild-type and *Ptnpf1* 2.8 chromatograms showing a 214 bp insertion on allele 1 and a 403 bp deletion on allele 2. C) *Ptnpf1* 2.9 mutation scheme and comparison between wild-type and *Ptnpf1* 2.9 chromatograms showing a 403 bp insertion on allele 1 and a 650 bp deletion on allele 2. gNPF1_a is underlined, including PAM sequence in the black box.
These two \textit{Ptnpf1} clones were further analysed, in order to predict the resulted mutated proteins. In fact, mutated gDNA sequences were translated using online Expasy Translate tools (\url{https://web.expasy.org/translate/}), and resulting amino acid sequences were compared with wild-type one. Mutations of both \textit{Ptnpf1} knock-out clones considered, result in frameshifts and/or stop codon insertion which block protein translation. This, as consequence, leads to the production of truncated proteins which could not properly fold and function.

In detail, in \textit{Ptnpf1} clone 2.8, the 403 bp deletion causes a frameshift from 240\textsuperscript{th} aa, with the consequent insertion of a stop codon which blocks protein translation at the level of the second transmembrane helix (TMH2). Similarly, the 214 bp insertion causes the introduction of additional residues starting from the 375\textsuperscript{th} aa and a subsequent stop codon that blocks the translation of the protein at the level of TMH6 (Fig. 3.17B). These modifications result in truncated proteins, which do not possess the key residues necessary for proper functioning.

In the same way, in \textit{Ptnpf1} clone 2.9, the 403 and 650 bp deletions cause frameshifts from 240\textsuperscript{th} and 24\textsuperscript{th} aa respectively, with a consequent stop codon introduction and the protein translation stop, at level of TMH5 in the first case and at the beginning of the protein in the second one (Fig. 3.17C). The resulting protein was truncated, unfolded and not properly functioning.

\textbf{Figure 3.17.} Structural models of \textit{P. tricornutum} NPF1 wild-type and knock-out. A) Structural model of PtNPF1 wild-type. B) Truncated protein model of PtNPF1 in the \textit{Ptnpf1} 2.8 strain resulting from the two mutations on the two alleles. C) Truncated protein models of PtNPF1 in \textit{Ptnpf1} 2.9 strain resulting from the two mutations on the two alleles.
3.3.6. Ptapt-Ptnpf2 knock-out mutants generation and screening

Similarly to PtNPF1 knock-out mutant generation, a proteolistic transformation was performed with the two gDNAs targeting PtAPT and the two gRNAs targeting PtNPF2 (Supplementary File S3.1).

After proteolistic transformation, more than 200 colonies positive to selection were obtained for each shot, and total 32 clones for Ptapt-Ptnpf2 knock-out were selected for further screening (Fig. 3.18A and Supplementary Table S3.1).

Firstly, insertion and/or deletions in PtNPF2 gene were confirmed by PCR: 40.6 % of PCR screened clones presents INDELs on the target gene (Fig. 3.18B). As to PtNPF1 (see before Results paragraph 3.3.5 “Ptapt-Ptnpf1 knock-out mutants generation and screening”), through electrophoresis gel it was possible to select more interesting strains for subsequent analyses, in particular Sanger sequencing was performed on these strains, in order to clarify INDELs length and to detect mutation-flanking SNPs. This last analysis was important to understand if single amplicon INDELs corresponded to biallelic homozygous or heterozygous mutations.

Figure 3.18. Selection and first PCR screening of Ptnpf2 proteolistic mutants. A) Positive C+ and negative C- control plates with non-transformed wild-type cells growing on F/2 medium without and with selection respectively, and four plates with transformed cells growing on selection. B) agarose gels showing the results of the PCR screening on PtAPT and PtNPF2 genes on wild-type
and transformed cells positive to selection, M represents the 100 bp marker (above and the 1Kb marker (below), while C- represents the blank.

Among the sequenced strains, *Ptnp2* clones 1.15 and 1.16 show biallelic heterozygous mutations, clearly defined by sequencing. In particular, *Ptnp2* clone 1.15 shows a 231 bp insertion and a 334 bp deletion on the two alleles respectively, while *Ptnp2* clone 1.16 presents a large deletion of 1181 bp and a smaller one, characterised by the insertion of an “A” nucleotide and the deletion of 5 bp (Figs. 3.19B-C and Supplementary File S3.3).
**Figure 3.19.** *P. tricornutum* wild-type and *Ptnpf2* mutants schemes and chromatograms showing Cas9 different cuts on the two *Ptnpf2* alleles. A) *PtNPF2* wild-type sequence showing gRNAs and primer couples used for sequencing. B) *ptnpf2* 1.15 mutation scheme and comparison between wild-type and *Ptnpf2* 1.15 chromatograms showing a 231 bp insertion on allele 1 and a 334 bp deletion on allele 2. C) *Ptnpf2* 1.16 mutation scheme and comparison between wild-type and *Ptnpf2* 1.16 chromatograms showing a 5 bp insertion and an “A” insertion on allele 1 and a 1181 bp deletion on allele 2. *gNPF2_a* is underlined, including PAM sequence in the black box.
These two strains were selected for further analyses, first of all to predict the resulting mutated proteins. Nucleotide sequences were translated using online tools, and resulting amino acid sequences were compared to wild-type to predict damages of PtNPF2 protein sequence and structure. In fact, mutations of both *Ptnpf2* clones result in frameshifts and/or stop codon insertion which blocks protein translation, resulting in the production of truncated proteins which could not properly fold and function.

In particular, *Ptnpf2* clones 1.15 deletion and insertion cause a sequence frameshift and a nucleotides insertion respectively, both starting from 166\(^{\text{th}}\) aa and both resulting in the introduction of a stop codon, which blocks the translation of the protein at the level of TMH4 (Fig. 3.20). As regards to *Ptnpf2* clones 1.16, the small deletion and the larger one cause a frameshift from 164\(^{\text{th}}\) aa and 161\(^{\text{st}}\) aa respectively, with the consequent introduction of a stop codon in both alleles, which blocks PtNPF2 translation at the level of TMH4 (Fig. 3.20). These modifications result in truncated proteins, which do not possess the key residues necessary for proper functioning.

**Figure 3.20.** Structural models of *P. tricornutum* NPF2 wild-type and knock-out. A) Structural model of PtNPF1 wild-type. B) Truncated protein models of PtNPF2 in the *Ptnpf2* 1.15 strain resulting from the two mutations on the two alleles. C) Truncated protein model of PtNPF2 in *Ptnpf2* 1.16 strain resulted from the two mutations on the two alleles.

*Ptnpf1* clones 2.8 and 2.9, together with *Ptnpf2* clones 1.15 and 1.16 were chosen for phenotypic characterisation of PtNPFs, described below in Chapters 4 and 5.
3.3.7. *Ptapt*-*Ptnpf1*-*Ptnpf2* double knock-out mutants generation and screening

The proteolistic protocol applied to generate single *PtNPFs* knock-out mutants was also optimised and then exploited to generate double *PtNPFs* knock-out strains, through the shot of RNPs containing Cas9 endonuclease and six gRNAs: two targeting *PtAPT* for selection, two targeting *PtNPF1* and two targeting *PtNPF2*.

After proteolistic transformations, more than 200 positive colonies were grown on selective plates (Fig. 3.21A). They were transferred on new plates and 77 of them were screened by PCR. Among these, 17 colonies show a mutation of only one *PtNPF* (mutations enough large to be visible from agarose gel, for example clones 1.23 and 1.31), while only two colonies present mutations of both *PtNPFs* genes, namely clones 1.29 and 1.30 (Fig. 3.21B and Supplementary Table S3.1).

**Figure 3.21.** Selection and first PCR screening of *Ptapt*-*Ptnpf1*-*Ptnpf2* proteolistic mutants (also called *Ptnpf1-2* knock-out mutants). A) Positive C+ and negative C- control plates with non-transformed wild-type cells growing on F/2 medium without and with selection respectively, and four plates with transformed cells growing on selection. B) agarose gels showing the results of the PCR screening on *PtNPF1* and *PtNPF2* genes on wild-type and transformed cells positive to selection, M represents the 1Kb marker and C- the two blanks.

These last two clones were sequenced to detect the mutation width. *Ptnpf1-2* knock-out strain 1.29 shows monoallelic mutations for both *PtNPF* genes: *PtNPF1* presents a 403 bp deletion on one allele, while the other one is not mutated, *PtNPF2* presents a 403 bp insertion on one allele and no mutations on the other one (Figs. 3.22A-B). In particular, as *Ptnpf1-2* knock-out strain 1.29 shows three amplicons on *PtNPF1* gene screening,
subclones were generated, but they present the same mutation pattern on agarose gel, probably due to heteroduplex formation during PCR reaction. On the other hand, Ptnpf1-2 knock-out strain 1.30 shows a heterozygous biallelic mutation for PtNPFl gene, with the insertion of 229 bp and 1 bp on the two alleles respectively, and a monoallelic mutation for PtNPF2, with one allele not mutated and the other one presenting a 581 bp deletion (Figs. 3.22C-D).

**Figure 3.22.** *P. tricornutum* wild-type and Ptnpf1-2 mutants chromatograms showing Cas9 different cut. Comparison between wild-type and Ptnpf1-2 knock-out strain 1.29 chromatograms showing A) a 403 bp deletion on only one PtNPFl allele and B) a 403 bp insertion on only one PtNPF2 allele. Comparison between wild-type and Ptnpf1-2 knock-out strain 1.30 chromatograms showing C) 229 bp and 1 bp insertions on PtNPFl alleles and D) a 581 bp deletion on only one PtNPF2 allele. gRNAs are underlined, including PAM sequence in the black boxes.
Although mutations identified in these two Ptntpf1-2 clones result in frameshift causing truncated protein, the presence of not mutated alleles did not allow to isolate clones with biallelic mutations for both PtNPFs.

3.3.8. Consecutive proteolistics to generate Ptapt-Ptnpf1-Ptnpf2 double knock-out mutants

As no Ptntpf1-2 double knock-out mutants with biallelic mutations of both PtNPF genes were obtained through previously described experiment, a different strategy was applied. Exploiting the other selective marker PtUMPS (ID: Phatr3_J11740) (Serif et al., 2018), conferring resistance to 5-FOA, proteolistic shots of PtUMPS and PtNPF2 gRNAs on Ptapt-Ptnpf1 knock-out background, and vice versa, were performed, following the protocol used to obtain PtNPFs single knock-out mutants.

First of all, 5-FOA drug efficacy was tested on wild-type P. tricornutum strains Pt1 and a Ptumps knock-out mutant previously generated by Serif et al. (2018). Increasing 5-FOA concentrations were tested on solid medium: 100, 300, 450, 600, 750 and 900 µg/ml (Fig. 3.23). While the original protocol used 100 µg/ml 5-FOA for selection, at this concentration some wild-type residual cells were observed on plates, so selective 5-FOA concentration was increased up to 600 µg/ml, to further reduce the risk of false positive colonies growing after transformations. This concentration was observed to be lethal for wild-type but not for Ptumps knock-out strain (Fig. 3.23). To note that 5 µg/ml uracil need to be added to Ptumps knock-out mutant cultures to allow their survival and growth, a consequence of the inactivation of the uracil biosynthesis cell system (Serif et al., 2018).
Figure 3.23. *P. tricornutum* Pt1 wild-type and Ptumps exposed to different 5-FOA concentrations. Cells plated on solid F/2 medium without silica, with uracil (URA) and with A) 100 µg/ml, B) 300 µg/ml, C) 450 µg/ml, D) 600 µg/ml, E) 750 µg/ml and F) 900 µg/ml 5-FOA.

Then, following protocol previously used to generate single *Ptnpf1* and *Ptnpf2* knock-out mutants, *PtUMPS* and *PtNPF2* gRNAs were assembled in RNP complexes and shot on *Ptapt*-Ptnpf1 knock-out background. No positive colonies growing on selective medium were obtained. Similarly, *PtUMPS* and *PtNPF1* gRNA were assembled in RNP complexes and shot on *Ptapt*-Ptnpf2 knock-out background, again without obtaining colonies growing on selective medium.

### 3.4. Discussion

*P. tricornutum* is one the main models among photosynthetic unicellular organisms. Techniques set up on it can be then adapted to other microalgae and applied for different purposes, from scientific research to technological applications. In fact, genetic
engineering on microalgae is spreading faster and faster in different fields of ecology, physiology and biotechnology, and the development of new methods solving each time the issues of the previous techniques is making these organisms more and more attractive, with the aim of shedding light on their ecological role or of exploiting them for more sustainable industrial applications.

The work of the DiaEdit (Development of genetic tools for the establishment of routine genome editing in the marine diatom *P. tricornutum*) consortium, for the development of genetic tools for the establishment of routine genome editing in the marine diatom *P. tricornutum* (Daboussi et al., 2014; Serif et al., 2018; Nymark et al., 2016) together with further functional studies (McCarthy et al., 2017; Moosburner et al., 2020; Hao et al., 2021) showed that genome editing in diatoms is feasible. To date different approaches have been used to generate transgenic lines and to screen the mutants, however progresses are at the beginning, compared to other systems (Kroth et al., 2018). So far genome editing has been performed in diatoms by integrating the genes encoding the different nucleases into the genome of the cells. Because of the big issues of this method (Stukenberg et al., 2018), new techniques were adapted to obtain higher transformation efficiencies, with lower frequency of off-target events (Karas et al., 2015; Serif et al., 2018).

In particular, proteolistic transformation for CRISPR/Cas9 application in diatoms is a DNA-free approach, that avoids random DNA integration into the genome, that drastically reduces off-target events, and that allows an antibiotic-free selection (Serif et al., 2018). The optimisation of this protocol, presented in this work, lead to further improvements, as it allows to increase the gRNAs:Cas9 ratio, while reducing the high costs derived by the big amount of Cas9 protein previously needed.

After successful reproduction of the original protocol in the laboratory, the adjustments in the RNP complexes assembly and in the selection procedure positively impacted method efficiency and costs. In particular, the use of a double drug concentration compared to the original protocol of Serif et al. (2018) allows to obtain less growing colonies on control plates, so decreasing the amount of false positive clones. The choice of increasing selection pressure can be balanced by the subsequent reduced screening efforts, which allows the faster identification of interesting mutated clones.

Moreover, while the original protocol proposed to assemble each gRNA with Cas9 separately and mix all the complexes only during the final steps before the shot, during
these experiments all the gRNAs needed for the transformation were mixed and together assemble with Cas9, at the beginning maintaining the same molar ratio between the different components. But the most important improvement was given by the further reduction of Cas9 amount used per transformation event. While the original protocol (Serif et al., 2018) proposed 8 µg Cas9 per transformation event, the optimised protocol here applied decreases to 1.5 µg the Cas9 quantity per shot. This improvement has been made possible by the new assembly procedure of gRNA-Cas9 complexes, but also by the reduction of the endonuclease quantity itself. In this way, the amount of Cas9 required has been reduced of more than five times, with two positive consequences: a further reduction of Cas9 presence in the cells, further lowering off-target cuts, and a reduction of transformation costs.

Decreasing five times the amount of Cas9, from 8 µg per shot of the Serif et al. (2018) protocol to 1.5 µg, while not changing the amount of total gRNAs, has also the interesting consequence of increasing the molar ratio between gRNAs and Cas9 protein. In fact, the original proteolistic protocol for diatoms (Serif et al., 2018), proposed a molar ratio gRNAs:Cas9 of 1:1, but in many recent works where CRISPR/Cas9 system has been applied to more established models, such as mice and human cells, the molar ratio gRNAs:Cas9 has been increased to 5:1 (Cheng et al., 2022) and more, reaching 10:1 (Min et al., 2019). Many genome engineering companies and platforms, such as Synthego, recommend to start with a molar ratio sgRNAs:Cas9 of 1:1 and testing ratios up to 9:1, as experiments showed that increasing the molar quantity of sgRNA relative to Cas9 increases the INDELs frequency. Through the optimised proteolistic protocol described in this Chapter, the molar ratio gRNAs:Cas9 has been increased to 6:1, while obtaining a big number of mutated clones.

Different experiments were performed to reach this point: the first experiment was done using only two gRNAs to test selection marker PtAPT and set up the modified protocol; then using four gRNA, so two for PtAPT and two for a PtNPF target gene. These shots allowed to generate Ptnpf1 and Ptnpf2 biallelic mutated strains which were subsequently analysed and used for phenotypic characterisation in the following Chapters 4 and 5. Then, experiments were performed using six gRNA: two for PtAPT, two for PtNPF1 and two for PtNPF2, with the aim of generating double PtNPFs knock-out mutants. Decreasing the amount of Cas9 did not impact the efficiency of transformation, as a
variable but big number of positive clones grew in selective plates and subsequent screening identified interesting mutations in them.

Anyway, no double *Ptnpf1-2* mutants with biallelic mutations for both *PtNPF* genes were identified, neither through this strategy nor through subsequent experiments of consecutive proteolitics. The failure to generate double *Ptnpf1-2* knock-out mutants could have different reasons, technical or physiological. As regards to the possible technical issues, the six-gRNAs allowed to obtain mutated strains but with a reduced efficiency in isolating biallelic mutations of all the genes of interest, so maybe more transformation experiments would be needed to generate expected mutated strains. Otherwise, the absence of resistant clones after consecutive proteolitics could be due to low efficiency of 5-FOA selection or could be caused by the excessive stress of altering four genes simultaneously. On the opposite, a physiological explanation could be the function of *PtNPFs* in regulating metabolic processes, so that the knock-out of both *PtNPF* genes could lead to a lethal phenotype. More experiments will be needed to verify one of these hypotheses.

The PCR screening of clones obtained from all the transformations, showed a different pattern of mutation. In particular, in most cases *P. tricornutum* alleles presented two different mutations, such as an insertion on one allele and a deletion on the other one, or two different deletions on the two alleles respectively. These occurred mutations were detectable through the first PCR screening because of their big length. This is mainly due to the position of the two gRNAs designed on each gene of interest, which possibly increases the chances to obtain large mutations, and allows a faster identification of interesting clones (Do et al., 2019; Acosta et al., 2018). Differently, isolated clones often need to be sequenced to see if there are mutations, or if mutations are biallelic, in particular when a single gRNA is designed, or two close gRNAs. So, the two gRNAs approach allows a fast screening and mutation identification (Acosta et al., 2018; Kraft et al., 2015), although clones would be anyway sequenced to verify the length of the mutations occurred.

This two-gRNAs system not only allows an easily detection of mutated strains, and in particular of biallelic mutations, but it also increases the possibility to isolate mutations which, even if not causing a frameshift and the insertion of a stop codon, could be large enough to substantially change the amino acid sequence and the protein structure to make it not functional (Wu et al., 2020). It can happen i) by inserting secondary structures which
avoid the correct protein folding, ii) by adding or removing important transmembrane helices (for example in transmembrane proteins such as the NPF transporters), iii) by removing or changing key amino acids involved in substrate recognition or signalling cascade, or iv) by modifying amino acids or entire domains necessary for conformational changes. Moreover, it has been observed that minor in-frame INDELs of non-critical residues can cause downregulation of the gene of interest, and could potentially be an alternative approach to alter metabolic pathways and to obtain interesting phenotypes (Wu et al., 2020).

After the PCR screening, mutations need to be sequenced in order to identify their width. This determination can be done manually on chromatograms, or using online tools providing a predictive genotype for each cell line that can be interpreted as homozygous, heterozygous, or mixed genotype. Among these tools, TIDER (http://tide.nki.nl/) (Brinkman et al., 2018; Moosburner et al., 2020) and ICE Analysis Tool (https://ice.synthego.com/#/ - Synthego Performance Analysis, ICE Analysis 2019, v3.0. Synthego). Here, the mutation study was performed manually, because of two main limitation of these online tools: i) here described proteolistic experiments were performed using two gRNAs per each gene of interest, while TIDER online tool does not allow to use both gRNAs as query (Moosburner et al., 2020); and ii) the big length of the obtained INDELs often required more sequenced fragment to be combined in a single consensus before align it with wild-type sequence, and both online tools do not still allow this step before the analysis, which then return uncomplete outputs.

On this Chapter, methods and screening of knock-out mutants obtained were presented, until the identification of interesting strains with biallelic mutations which could be further analysed. Once a mutation has been identified in the genome, there are different ways to check if the mutation affects the final protein structure or functioning.

In fact, INDELs resulting in a frameshift generate a premature termination codon (PTC) (Ran et al., 2013). Owing to the truncated coding sequence, the corresponding mRNA is usually recognised by the nonsense-mediated mRNA decay (NMD) pathway, which degrades transcripts harbouring PTCs (Karousis et al., 2016; Reber et al., 2018), resulting in a de facto knock-out. However, NMD is not able to target all containing a PTC with the same degradation efficiency. This is due to the position of the PTC within the transcript and the sequence composition around the termination codon (Reber et al., 2018), which strongly influence the sensitivity of the NMD for a given transcript. Thus,
different transcripts can escape NMD and in some cases the corresponding truncated proteins are detectable. Consequently, transcripts harbouring frameshift mutations mediated by CRISPR/Cas9 may give rise to C-terminally truncated proteins, which could have residual or dominant negative functions (Reber et al., 2018).

An efficient approach is to directly check the protein: if antibodies are present, correct protein folding can be confirmed through Western blots. However, this way is not always a viable option, because residual truncated proteins could be detected by antibodies. Moreover, in this specific case, clones here generated own heterozygous biallelic mutations, which can lead to a problematic interpretation of the results, in particular in case of a deletion on one allele and an insertion in the other one. Finally, no antibodies are still available for PtNPF genes of interest, so that protein studies could require too much time and efforts.

In this Chapter, the application and the optimisation of a proteolistic protocol to generate P. tricornutum knock-out mutants through the CRISPR/Cas9 system has been described, but much progress is being continuously made in diatom genome editing.

Multiplex knock-out is the new frontier that researchers are chasing in recent years. It means the simultaneous cut of different target genes, and can be performed in two different ways: i) using many gRNAs targeting specific region on the genes of interest (Hao et al., 2021; Serif et al., 2018), similarly to experiments here performed to generate double Ptnpf1-2 knock-out mutants, or ii) designing two sgRNAs that could potentially target all the genes belonging to same gene family (Sharma et al., 2021a). In fact, when working with gene families with functional redundancy, knocking out multiple genes can be required to generate a phenotype. In this context, it is possible to exploit the known tolerance of Cas9 for mismatches between the gRNA and the target site to simultaneously introduce INDELs in multiple homologous genes (Sharma et al., 2021a; Kleinstiver et al., 2016).

Another new frontier of CRISPR/Cas9 application in diatoms has been recently explored by Nam et al. (2022), who developed a single vector CRISPR/Cas9 guided GFP knock-in strategy in the diatom T. pseudonana. This system enables precise and scarless knock-in of GFP at the endogenous genomic location to create GFP fusion proteins under their native cis and trans regulatory elements, providing at the same time CRISPR/Cas9 gene editing and robust fluorescent protein tagging (Nam et al., 2022).
Alternatives to CRISPR/Cas9 systems are also being developed as new frontiers of genome editing. Among them, the site-specific recombinase (SSR) approach in diatoms. It involves an integrase (Int), which catalyses the integration (between the external \textit{attP} and the host \textit{attB} sites) and the excision (between the recombinant \textit{attR} and \textit{attL} sites) of the sequence into and out of the chromosome of its host, through site-specific recombination reactions (Kolot et al., 2015; Kroth et al., 2018). Native Int \textit{attB} sites, conserved in many plant and animal organisms, were identified on the chromosomes of \textit{P. tricornutum}, and may be used for genome manipulations via Int-catalysed recombinase-mediated cassette exchange reactions (Kroth et al., 2018).

Genome editing technologies are revolutionising the way to modify DNA and alter gene expression in cells and organisms. In this context, RNP delivery is a new method avoiding the bigger issues of existing CRISPR/Cas9 knock-out techniques. Its simplicity and ease of implementation, as it eliminates all subcloning steps, saves considerable time relative to traditional genome-editing approaches (Serif et al., 2018). For this reason and because \textit{UMPS} and \textit{APT} markers are well conserved within the microalgae phylogenetic tree and among other eukaryotic groups, proteolistic method is extendable to many other organisms.
Chapter 4: *P. tricornutum* NPF1 functional characterisation
Abstract

Nutrient availability in the environment is an essential factor influencing life and shaping ecosystems. The need to cope with nutrient fluctuations and often their scarcity is a key driver of evolution, as it pushes uni- and multicellular organisms to develop ever more efficient strategies to store nutrients. The finely regulated balance between circulating nutrient levels and intracellular stored nutrients allows to maintain the equilibrium with the external environment. However molecular mechanisms controlling dynamics of nutrient storage and mobilisation are still poorly known in many organisms, among which diatoms.

Diatoms are one of the most successful phytoplanktonic groups dealing with strong nutrient fluctuations in the ocean. One of diatom traits explaining their capability to outcompete other microorganisms in variable conditions is the presence of the vacuole, which can store and make again available very high concentrations of nutrients, in particular nitrate.

Here, new insights were provided on *P. tricornutum* Low-affinity Nitrate Transporter NPF1, a predicted low-affinity nitrate transporter localised on the vacuole membrane, putatively involved in nitrogen reallocation inside the cell and in nitrogen fluxes regulation.
4.1 Introduction

Nutrient scarcity has certainly acted as selective pressure for all living organisms so that, in scenarios of intermittent availability, the major adaptive solution is represented by nutrient storage. In fact, a flexible and highly regulated balance between circulating nutrient levels and stored nutrients allows to cope with nutrient variability in the ocean (Efeyan et al., 2015). Among strategies implemented by organisms to cope with nutrient limitation and store important resources, vacuoles are one of the most successful.

The vacuoles are membrane-bound organelles that contain inorganic ions and organic compounds, important compartments of many eukaryotic cells. Their morphology and biochemistry, as well as the materials they store, vary greatly depending on the cell type and stage of plant development (Noguchi and Hayashi, 2014). In fact, vacuoles can be of different types, depending on the function they perform in different organisms.

For example, protozoa have specialised acidic vacuoles called food vacuoles, for ingesting and recycling organic compounds (Chugh et al., 2013), and contractile vacuoles responsible for osmoregulation (Komsic-Buchmann et al., 2014). In higher plants, the two most studied vacuoles are the central (or lytic) one and the protein storage vacuole. The first, large in size and mainly present in leaves, is involved in osmoregulation, structure maintenance and molecular degradation and storage, in order to maintain a balance between biogenesis and degradation of many substances. While the protein storage vacuole, smaller in size and responsible for secondary metabolites storage, accumulates reserve proteins or soluble carbohydrates (Martinoia et al., 2006; Schreiber et al., 2017). Other kinds of vacuoles include those involved in interactions with other organisms, such as parasitophorous and symbiontophorous vacuoles (Kostygov et al., 2016; Schreiber et al., 2017).

The vacuolar membrane, called tonoplast, contains transport proteins with different functions. They can maintain cytoplasm homeostasis such as proton (H⁺) pumps (vacuolar H⁺-ATPase and H⁺-pyrophosphatase) which stabilise cytoplasmic pH; they can facilitate osmotic water flow across membranes, such as aquaporins which can increase the permeability of the osmotic membrane to water approximately 100-fold compared to that of a pure phospholipid bilayer (Komsic-Buchmann et al., 2014); they can selectively transport ions such as Cl⁻, Na⁺, Ca²⁺, K⁺, nitrate (NO₃⁻) and phosphate.
(PO₄⁻) (Noguchi and Hayashi, 2014); or they can transport bigger molecules such as malate and glutamate (Martinoia et al., 2006).

In particular, plant vacuoles are identified as the major NO₃⁻ storage pools and contain up to 90% of the total cellular NO₃⁻ (He et al., 2017), helping to maintain cell metabolism when external N supply is limited. However, vacuolar NO₃⁻ is not readily accessible to the cytosolic Nitrate Reductase (NR), thus it has to be reallocated for metabolic use when necessary: this is possible thanks to the NO₃⁻/H⁺ antiport machinery and the NO₃⁻/H⁺ symport system (He et al., 2017). In higher plants, these machineries include the chloride channel CLC, the high-affinity NO₃⁻ transporter NRT2.7 and the two low-affinity NO₃⁻ transporters NPF (NRT1/PTR Family) PTR4, PTR6 well studied and characterised in Arabidopsis thaliana (Fan et al., 2017).

Diatoms hold vacuolar-like structures, surrounded by only one membrane similarly to higher plants. However, these large structures have been poorly studied, so far, although knowledge of the vacuolar structure and function is essential for understanding physiology of nutrition and stress tolerance of microalgae.

In the last years the interest in vacuolar functions in the microalgal cell increased thanks to recent progresses in biotechnology (Shebanova et al., 2017): diverse high-value compounds could be synthesised by microalgal cells under stress conditions, particularly by nutrient limitation, in which vacuoles play an essential role (Pittman et al., 2011).

The main explanation attributed to the development of vacuoles by diatoms is based on the assumptions on their ecological niche. In fact, diatoms are one of the most successful groups in the plankton, and one of the main reasons is that they have managed to thrive in so many different habitats (Kooistra et al., 2007). In particular, diatoms are well adapted to oligotrophic environments, as they are naturally equipped to absorb and store inside the cell a very big amount of nutrients, against future nutrient shortage (Shebanova et al., 2017).

However, similarly to higher plants, accumulation of these nutrients (mainly P and N) in directly bioavailable form is not possible: high levels of polyphosphate (PolyP) can be toxic in the cell when accumulated in the cytoplasm (Gerasimaitė et al., 2014; Shebanova et al., 2017). Moreover, many diatoms take up and store NO₃⁻ intracellularly in concentrations of up to 100 mM, exceeding ambient NO₃⁻ concentrations by several orders of magnitude (Kamp et al., 2011), which is toxic if directly available by the cell.
Nutrients, such as PO$_4^-$ and NO$_3^-$, uptake and accumulation in vacuoles occur during favourable periods and conditions when they exceed in the environment, a phenomenon also called luxury uptake (Behrenfeld et al., 2021). While nutrients release in cytosol occur in different unfavourable conditions, making them bioavailable for cell metabolism. For example, in dark and anoxic conditions, some diatoms can consume 84-87% of their intracellular NO$_3^-$ pool within one day through the Dissimilatory NO$_3^-$ Reduction to Ammonium (NH$_4^+$) (DNRA), which is an anaerobic respiration process used by many microbes to enter a resting stage for long-term survival (Kamp et al., 2011). Moreover, NO$_3^-$ internal reservoir in vacuole could support growth in N starvation condition (Wang et al., 2019).

As the role of vacuoles in diatoms is still little known, also the function of many transporters localised on the tonoplast needs to be characterised. These proteins are often key players in mechanisms of molecules storage in and release from vacuolar environment, allowing exchanges between different cellular compartments. In this way, cells have the required resources for metabolism available, and maintain a favourable intracellular environment by storing excess resources such as luxury nutrients (Behrenfeld et al., 2021).

Recently, some vacuolar proteins have been identified and functionally characterised in diatoms, thank to genomic and genetic developments in the last years. A study of Huang et al. (2016) presented two 1,6-β-transglycosylases localised on the tonoplast and responsible of synthesising and transport chryolaminarin, which is the main storage carbohydrate of diatoms. In another recent research, Dell’Aquila et al. (2020) showed the Vacuolar Transport Chaperone (VTC) complex to relocate from the nuclear envelope/endoplasmic reticulum to the vacuole under conditions of PO$_4^-$ limitation, hypothesising that this protein could have a role in P influx/efflux in diatom vacuole.

However, many other of the vacuolar transporters are still considered orphans.

The diatom low-affinity nitrate transporters NPFs (diNPFs) have proved to be a very diverse and interesting class of transporters. Interestingly, phylogenomic analyses revealed that they diverge into two clades, and 25% of diNPFs, all belonging to the plant-like Clade II, were predicted to the tonoplast (Santin et al., 2021b).

Here, new insights on diNPFs Clade II and on their role within cellular metabolism are provided. In order to characterise them for the first time, *Phaeodactylum tricornutum* was
chosen as model organism. This pennate diatom is well-characterised in terms of functional genomics, and for this species many innovative genomic and genetic resources became recently available such as the CRISPR/Cas9 genome editing technology (Nymark et al., 2016; Serif et al., 2018; Russo et al., 2022; Falciatore et al., 2020). *P. tricornutum* owns two NPFs, of which PtNPF1 belongs to Clade II.

Here, by molecular and functional genetic analyses, we unveil that the *PtNPF1* gene could play a pivotal role in regulating the luxury N uptake, helping to explain the unique ability and/or plasticity of these successful microalgae to store and reallocate intracellular N resources.

### 4.2. Materials and Methods

#### 4.2.1. Diatom cultures

An axenic culture of *Phaeodactylum tricornutum* Bohlin CCMP 632 was obtained from the Provasoli-Guillard National Centre for Culture of Marine Phytoplankton. The culture was maintained in autoclaved 0.22 μm filtered F/2 medium without silica (Guillard, 1975), temperature was controlled at 18°C and white fluorescent lights were used at a 12:12 light dark cycle with a photon flux of 90 μmol photons m$^{-2}$ s$^{-1}$. All cultures in this study were grown in the same way under these conditions, except for specific growth curves experiments described below.

#### 4.2.2. In silico prediction of PtNPF1 vacuolar targeting

To study in more detail the subcellular localisation of PtNPF1, in addition to the previous in silico analyses carried out using LocTree, WolfPSort and SignalP (see paragraph 2.3.8 “Predicted subcellular localisation” in Chapter 2), further analyses were performed. In particular, FIMO version 5.4.1 ([https://meme-suite.org/meme/tools/fimo](https://meme-suite.org/meme/tools/fimo)) was used to search in the PtNPF1 amino acid sequence, the conserved motif [D/E]xxxL[L/I] for vacuole localisation previously identified by Bonifacino and Traub (2003).

#### 4.2.3. PtNPF1 overexpression vectors

*P. tricornutum* cells were transformed with a two-vector system: an overexpression vector containing the *PtNPF1* gene of interest fused with a gene encoding for a fluorescent protein and under the control of a strong promoter, and a vector with a gene that confers
resistance to the antibiotic phleomycin, to provide selection for screening successful transformants (Falciatore et al., 1999; Zaslavskaya et al., 2001).

*PtNPF1* gene (Phatr3_J47148) was fused to the *YFP* gene at its 3’ end or with the *GFP* gene at its 5’ end.

In detail, in the first case, the expression was under the regulation of the Light Harvesting Complex *Lhcf2* (before called fucoxanthin-chlorophyll binding protein-B *FcpB*) strong promoter, *Lhcf2p*, and the Light Harvesting Complex *Lhcf1* (before called FcpA) terminator, *Lhcf1t*. Cloning was performed using the Gibson Assembly reaction (Gibson et al., 2010, 2009) and NEBuilder Assembly Tool (http://nebuilder.neb.com/) as suggested by manufacturer’s instructions. The pKS-FcpBpAt-C-EYFP vector (Siaut et al., 2007) backbone of 4357 bp containing promoter, 3’ *YFP* gene and terminator was amplified using primers *Lhcf2p*YPF_for and *Lhcf2p*YFP_rev, full-length *PtNPF1* sequence of 2328 bp was amplified from *P. tricornutum* genomic DNA using primers *Lhcf2p_PtNPF1_for* and *PtNPF1_YFP_rev*, allowing the subsequent insertion of the *PtNPF1* gene without stop-codon between promoter and *YFP* gene (Table 4.1). Q5™ High-Fidelity DNA Polymerase was used to amplify the fragments, using the thermal profile [98°C for 30 sec, (98°C for 10 sec, 67°C for 40 sec, 72°C for 2 min) ×35 cycles, 72°C for 2 min]. This construct of 6682 bp, was called Lhcf2p-PtNPF1-YFP-Lhcf1t vector.

While, for the *GFP* gene at 5’ end of the *PtNPF1* gene, the expression was under the control of the *Pseudo-nitzschia multistriata* histone *H4* constitutive promoter (*PmH4p*). In particular, the *PmH4pH4N-GFP* plasmid (Sabatino et al., 2015) was modified through Gibson Assembly reaction (Gibson et al., 2010, 2009): plasmid backbone of 4457 bp containing promoter, 5’ *GFP* gene without stop-codon and terminator, was amplified using primers *PmH4pGFP_for* and *PmH4pGFP_rev*, and it was assembled with the full-length *PtNPF1* sequence previously amplified using *GFP_PtNPF1_for* and *PtNPF1_Lhcf1t_rev* (Table 4.1), so inserting the *PtNPF1* gene between *GFP* gene and *Lhcf1* terminator and creating the 6785 bp *PmH4p-GFP-PtNPF1-Lhcf1t* vector. Also in this case, Q5™ High-Fidelity DNA Polymerase was used to amplify the fragments, using the thermal profile [98°C for 30 sec, (98°C for 10 sec, 67°C for 40 sec, 72°C for 1 min and 30 sec) ×35 cycles, 72°C for 2 min].
Table 4.1. List of primers used for plasmid construction.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lhc2pYPF_for</td>
<td>5’-GAATTCGATATCAAGCTTATCGAT-3’</td>
</tr>
<tr>
<td>Lhc2pYFP_rev</td>
<td>5’-GAATTCATGGTGAGCAAGGCG-3’</td>
</tr>
<tr>
<td>PmH4pGFP_for</td>
<td>5’-GCGGCCGCAACAACTACC-3’</td>
</tr>
<tr>
<td>PmH4pGFP_rev</td>
<td>5’-AGGCCCTCTTGTACAGCTCG-3’</td>
</tr>
<tr>
<td>Lhc2p_PtNPF1_for</td>
<td>5’-atcgataagcttgatatagttgcagTGAGCTCCAGTGGAGAG-3’</td>
</tr>
<tr>
<td>PtNPF1_YFP_rev</td>
<td>5’-cccttgctcaccatgaattcGGAGGACGTTGCTCGCAAC-3’</td>
</tr>
<tr>
<td>GFP_PtNPF1_for</td>
<td>5’-acgagctgtaacaggtggctATGACGACTCCAGTGGAG-3’</td>
</tr>
<tr>
<td>PtNPF1_Lhc1t_rev</td>
<td>5’-gaggttagttgtaggccccgCTAGAGGACGTTGCTCG-3’</td>
</tr>
</tbody>
</table>

To perform the Gibson assembly, a total of 0.03 pmols of the backbone and 0.07 pmols of the PtNPF1 fragment were assembled into a single vector, at 50°C for 1 hour. Transformation was carried out in the One Shot™ TOP10 Chemically Competent E. coli strain (Invitrogen) through heat shock: 2 μl of the Gibson assembly were added to 100 μl of cells, put on ice for 10 minutes, exposed to shock at 42°C for 30 seconds, put on ice for 2 minutes. 1 ml of LB medium was added to cells and put for 1 hour at 37°C. Then, cells were plated on LB solid medium containing 100 μg/ml ampicillin (Sambrook et al., 1989) and put at 37°C overnight. Diagnostic PCR was carried out on clones growing on selective medium, plasmidic DNA was extracted from positive clones, using the GenElute™ Plasmid Maxiprep Kit (Sigma-Aldrich) according to manufacturer’s protocol, and then sequenced.

To confer resistance to the antibiotic Phleomycin, a second vector with the Sh-ble gene under the control of the Light Harvesting Complex Lhc6 (before called fucoxanthin-chlorophyll binding protein-F FcpF) promoter, Lhc6p, vector named Lhc6p-Sh-Ble-Lhc1t (Falciatore et al., 1999), was co-transformed with one of the PtNPF1-YFP or GFP-PtNPF1 expression vector, respectively.

4.2.4. Biolistic transformation of *P. tricornutum*

The transformation of the axenic CCMP632 strain of *P. tricornutum* Bohlin was performed by microparticle bombardment using the Biolistic PDS-1000/HE Particle Delivery System (Bio-Rad), through previously described biolistic methods (Falciatore et al., 1999). Briefly, two-vector transformations, tungsten M-17 nanoparticles (1.1 μm
diameter, Bio-Rad) were coated with 3 µg of one of previously described PtNPF1 overexpressing vectors (Lhcf2p-PtNPF1-YFP-Lhcf1t or PmH4p-GFP-PtNPF1-Lhcf1t) and 3 µg of the FcpFp-Sh-Ble-Lhcf1t plasmid, using the CaCl₂-spermidine method: to the M17-plasmids mix, 1.25 M CaCl₂ and 20 µM spermidine were progressively added, while continuously vortexing. After 3 minutes vortex, the mix was centrifuged at maximum speed for 10 seconds, washed with absolute ethanol, centrifuged again and resuspended in 60 µl absolute ethanol, enough for five transformations.

10 µl of the mix were shot into 5 · 10⁷ cells plated on 50% F/2 medium without silica. In particular, this solid medium consisted of 50% filtered seawater, 50% Milli-Q water and 1% agar supplemented with nutrient stocks commonly used for F/2 medium (Guillard, 1975). The biolistic shot was performed as described in the paragraph 3.2.7 “Transformation and replating” in Chapter 3, following Falciatore et al. (1999). After growing for 2 days on nonselective media, plated cells were transferred to classic F/2 agar plates without silica, with 1% agar and 50 µg/ml of phleomycin for selection. Colonies, positive for phleomycin resistance, appeared between 20 and 30 days.

4.2.5. PCR screening on transformed cells overexpressing PtNPF1

PCR was performed on genomic DNA from transformants and wild-type cells, picked from selective and control plates. Cells were then resuspended in 20 µL of a Lysis Buffer consisting of 1% (v/v) Triton X, 20 mM Tris-HCl (pH 8) and 2 mM EDTA (pH 8) in Milli-Q water, vortexed, put on ice for 15 minutes and then at 85°C for 10 minutes, as described in Daboussi et al. (2014). The diluted solution was then used for PCR screening.

To analyse NPF1-YFP cassette integration in genomic DNA, the thermal profile used was [95°C for 1 min, (95°C for 20 sec, 58°C for 20 sec, 72°C for 2 min and 30 sec) ×35 cycles, 72°C for 2 min]. The primers used are listed in Table 4.2: Lhcf2p_for PtNPF1_exp_rev amplified a fragment of 1994 bp, PtNPF1_exp_for and GFP_rev amplified a fragment of 1336 bp.

While to analyse GFP-NPF1 cassette integration in genomic DNA, the thermal profile used was [95°C for 1 min, (95°C for 20 sec, 62°C for 20 sec, 72°C for 2 min) ×35 cycles, 72°C for 2 min] with primers PmH4p1_for and GFP_rev (fragment 1319 bp), and [95°C for 1 min, (95°C for 20 sec, 58°C for 20 sec, 72°C for 2 min and 30 sec) ×35 cycles, 72°C for 2 min] with primers GFP_for and PtNPF1_exp_rev (fragment 2014 bp) (Table 4.2). Amplification fragments were analysed on agarose gel.
<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lhcf2p_for</td>
<td>5’-CGCCGTAACAGCAAATCCT-3’</td>
</tr>
<tr>
<td>PmH4p1_for</td>
<td>5’-AACAAAGCTGGGTACCGGC-3’</td>
</tr>
<tr>
<td>PtNPF1_exp_for</td>
<td>5’-CTACGAAGTCGCCTTTACCG-3’</td>
</tr>
<tr>
<td>PtNPF1_exp_rev</td>
<td>5’-ATCTTCCAACCGCGGTATAC-3’</td>
</tr>
<tr>
<td>GFP_for</td>
<td>5’-CCAGCAGAACACCCCCAT-3’</td>
</tr>
<tr>
<td>GFP_rev</td>
<td>5’-AECTCCAGCAGGACCATGTG-3’</td>
</tr>
</tbody>
</table>

### 4.2.6. Confocal Microscopy

The vacuoles of wild-type cells, PtNPF1 overexpressing strains and Ptnpf1 knock-out mutants (see paragraph 3.3.5 “Ptapt-Ptnpf1 knock-out mutants generation and screening” in Chapter 3) were observed with a laser-scanning confocal microscope Leica SP8 X. All images were collected using the HC PL APO CS2 63X/1.20 water objective.

Chlorophyll a autofluorescence was excited at 554 nm and detected at 650-741 nm.

For the visualisation of the structure of vacuoles in *P. tricornutum* wild-type and Ptnpf1 knock-out mutants, a green fluorescent tonoplast tracker, the Yeast Vacuole Membrane Marker MDY-64 (Invitrogen) was used with a final concentration of 2 μM (Huang et al., 2016): 1 μl of 1 mM MDY-64 was diluted into 500 μl liquid samples (containing about $10^6$ cells) of *P. tricornutum* in F/2 medium, and then incubated at room temperature for 2 minutes. MDY-64 was excited at 451 nm with a 65-mw Argon laser and detected at 490-562 nm of emission (Supplementary Table S4.1). YFP and GFP fluorescence emission signal, in PtNPF1 overexpressing strains, was detected at 500-562 nm by excitation at 488 nm.

### 4.2.7. Vacuole morphology measurements

In order to observe any effect of N starvation on the *P. tricornutum* vacuole, wild-type cells were grown in 882 μM and 50 μM NaNO₃, respectively, and images were taken through the confocal microscope after four days. Then, in order to study the possible involvement of PtNPF1 in vacuolar morphology changings, *P. tricornutum* wild-type, PtNPF1 overexpressing strains and Ptnpf1 knock-out mutants were grown in 882 μM and
50 µM NaNO$_3$, respectively, and images were taken through the confocal microscope after 2, 4 and 7 days, as described in the previous paragraph. Each condition was set up in triplicate. In detail, to visualise the vacuoles, wild-type and $Ptnpf1$ knock-out mutants were stained with the MDY-64 vacuole tracker, while YFP fluorescence of fusion proteins was exploited in $PtNPFl$ overexpressing strains.

For each cell, the number of vacuoles was counted, and their area was measured using ImageJ (Fiji) (Schindelin et al., 2012). The ratio between the total area of vacuoles and the total area of the cell was then calculated. In the same way, the ratio between the total area of the chloroplast and the entire cell was measured.

4.2.8. Growth curves and N starvation and repletion experiment

$P. tricornutum$ wild-type, $PtNPFl$-YFP overexpressing strains 1OE 4 and 1OE 9, and $Ptnpf1$ knock-out lines 1KO 2.8 and 1KO 2.9 were grown axenically in F/2 medium without silica (Guillard, 1975) to mid-exponential phase and then transferred to different medium conditions to perform growth curve experiments. A knock-out mutant for a different gene, generated with the same proteolistic protocol (2KO 1.15), was used as additional control strain. All growth curves were performed at 18°C and white light was provided at a 12:12 light dark cycle with a photon flux of 90 µmol photons m$^{-2}$ s$^{-1}$.

Cells were grown in triplicate in vented cap flasks at different NO$_3^-$ concentrations and different N sources for 10 days: F/2 medium without silica with 882 µM NaNO$_3$ set as reference condition, 50 µM NaNO$_3$ as N starvation condition, and 882 µM NH$_4$Cl or 882 µM urea as the alternate N sources. Cell concentration was evaluated using a Malassez cell counting chamber and by the in vivo chlorophyll a fluorescence, a proxy for growth (Supplementary Fig. S4.1), measured through a multifunctional monochromator-based microplate reader (Infinite™ M1000 Pro; Tecan) with excitation and emission wavelengths set at 662 and 685 nm, respectively (Mandalakis et al., 2017; Kalaji et al., 2014; Ni et al., 2019).

For the N starvation and repletion experiment, $P. tricornutum$ wild-type and mutant strains were diluted to 2 · 10$^5$ cells/ml and transferred in triplicate in 500 ml F/2 medium without silica with 50 µM NaNO$_3$ as N source for four days and then the starved cultures were resupplied providing NaNO$_3$ to a final concentration of 882 µM. The same experiment was performed by repleting cells with 882 µM NH$_4$Cl or 882 µM urea, after four days of N starvation. Moreover, cells were starved with no P for four days and then
resupplied with 36 µM NaH$_2$PO$_4$ · H$_2$O. Each condition was set up in triplicate. The cell concentration was evaluated through the flow cytometer BD FACSVerse™ (BD Biosciences) and through *in vivo* chlorophyll *a* fluorescence.

For gene expression analyses, cells were collected at four time-points: at the mid-exponential phase on 882 µM NaNO$_3$ at the beginning of the experiment (T0), after four days of growth on 50 µM NaNO$_3$ (N starvation – T4), after one (T5) and three (T7) days after NO$_3^-$ repletion. All cultures were incubated in diel light (see above) at 18°C.

**Figure 4.1.** Pipeline of the N starvation and repletion experiment, including the list of *P. tricornutum* strains used, the time points analysed and the measurements performed. T0 indicates the beginning of the experiment, before the N starvation beginning, T4 indicates N starvation reached after four days in 50 µM NaNO$_3$, T5 indicates the day after repletion with 882 µM NaNO$_3$ (5 days from the beginning of the experiment) and T7 indicates three days after N repletion (7 days from the beginning of the experiment). The three flasks represent three biological replicates for each strain.

**4.2.9. Extra- and intra-cellular NO$_3^-$ content analyses**

To analyse extra- and intra-cellular NO$_3^-$ pools, samples of 4 · 10$^7$ cells (different volumes depending on the cells concentration) were harvested onto GF/F filters. Prior to filtration, each filter was washed with N-free F/2 medium made with artificial seawater (Dortch, 1982). Filters were saved in a microcentrifuge tube, covered with 1.5 ml of N-free F/2 medium, and quick-frozen in liquid nitrogen. The resulting volume of filtrate medium were collected in 15-mL tubes and stored in the dark at 4°C. Cell contents were extracted
according to McCarthy et al. (2017): the microcentrifuge tubes were heated in a 100°C bath for 10 minutes, then cooled on ice for 5 minutes, vortexed and centrifuged for 10 minutes at 10 000 g at 4°C. Filters were washed with the same supernatant and then discarded. Cell extracts were centrifuged again briefly, the clarified extract was transferred to a clean 1.5 ml tube and stored at 4°C. NO$\text{}_3^-$ in the extract and in the filtrate was measured by UV spectrophotometry (Cary 100 UV-Vis spectrophotometer - Varian Inc., single-beam spectrophotometer) at a wavelength of 220 nm (Collos et al., 1999; McCarthy et al., 2017). Outputs were correlated to a NO$\text{}_3^-$ standard curve (NaNO$\text{}_3$ concentrations from 5 to 800 µM in N-free F/2 medium) established at the beginning of the experiment and rerun prior to each use of the Cary 100.

4.2.10. Statistical analyses

To determine whether there were significant differences in morphological features, growth, intra- and extracellular NO$\text{}_3^-$ content between the different strains and over time, a two-way ANOVA with Tukey’s Multiple Comparison Test was performed, and statistics for time, strain, and interaction were recorded. Statistics was performed using GraphPad Prism version 6.0 and values were defined significantly different when $p < 0.05$.

4.2.11. RNA extraction and qPCR analysis

RNA extraction and quantitative Real-Time reverse-transcription Polymerase Chain Reaction PCR (qRT-PCR or qPCR) were performed as described in Russo et al. (2015). For RNA extraction, $10^8$ cells from exponential phase cultures were collected by centrifuging for 10 minutes at 4000 g at 4 °C. After being washed with PBS (1X) to completely eliminate sea salts, pellets were stored at -80°C until RNA extraction. So, total RNA was isolated using 1.5 ml Trizol™ (Invitrogen), as previously described in paragraph 2.2.6 “RNA extraction, primer design and qPCR” in Chapter 2.

RNA quantity was determined using Qubit™ 2.0 Fluorometer (Life Technologies), while RNA quality was checked by gel electrophoresis. 200 ng of RNA were retro-transcribed using the PrimeScript™ RT Reagent Kit (Perfect Real Time) (TakaraBio). To assess cDNA quality and absence of gDNA contamination, Reverse Transcriptase (RT) -PCRs were run with control PtNOA gene (Nitric Oxide Associated protein, Phat3_J40200) using intron spanning primers PtNOA_for and PtNOA_rev.
For the qPCR, 1 μl of a 1:2 dilution of cDNA was used as template to amplify the transcripts using 0.4 μM final concentration of the primers. The *RPS* gene (Ribosomal Protein Small Subunit 30S, Phatr3_J10847) was used as reference gene for qPCR data normalization (Siaut et al., 2007). All the genes and primers used for qPCR analysis are listed in Table 4.3.

<table>
<thead>
<tr>
<th>Target gene</th>
<th>Gene ID</th>
<th>Primer name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PtNOA</td>
<td>40200</td>
<td>PtNOA_exp_for</td>
<td>5’-CAGTTACTGACCCCGAAGA-3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PtNOA_exp_rev</td>
<td>5’-AACGCACCTTTCCGAGAGA-3’</td>
</tr>
<tr>
<td>PtRPS</td>
<td>10847</td>
<td>PtRPS_for</td>
<td>5’-GTGCAAGAGACCGAGATACC-3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PtRPS_rev</td>
<td>5’-CGAAGTCAACGGGAAACCA-3’</td>
</tr>
<tr>
<td>PtNPFs</td>
<td>47148</td>
<td>PtNPF1_exp_for</td>
<td>5’-CTACGAAGTGCCCTTACCCG-3’</td>
</tr>
<tr>
<td></td>
<td>47218</td>
<td>PtNPF1_exp_rev</td>
<td>5’-ATCTTCCACCGGTGATAC-3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PtNPF2_exp_for</td>
<td>5’-TTACGTGATGGCTTGATCCA-3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PtNPF2_exp_rev</td>
<td>5’-GGTCCGGGTTATACAGAGA-3’</td>
</tr>
<tr>
<td>PtNRT2</td>
<td>26029</td>
<td>Pt26029_exp_for</td>
<td>5’-TGGGATATATTTCTTTACGTA-3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pt26029_exp_rev</td>
<td>5’-GCATGCGCTTCTTCAAGTCAG-3’</td>
</tr>
<tr>
<td></td>
<td>54101</td>
<td>Pt54101_exp_for</td>
<td>5’-GGAGTGCCGGGAAATTTTT-3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pt54101_exp_rev</td>
<td>5’-ATGACAGACGGAAGACCGCC-3’</td>
</tr>
<tr>
<td></td>
<td>54560</td>
<td>P54560_exp_for</td>
<td>5’-GCATCGAGCGCAAATCTTTT-3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P54560_exp_rev</td>
<td>5’-GCTACCGGACTTTTTGTA-3’</td>
</tr>
<tr>
<td></td>
<td>40691</td>
<td>Pt40691_exp_for</td>
<td>5’-GCCACCAGTCTATTTGCGA-3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pt40691_exp_rev</td>
<td>5’-TATAGCCAGGTCTACGCTCG-3’</td>
</tr>
<tr>
<td></td>
<td>2032</td>
<td>Pt2032_exp_for</td>
<td>5’-TGACCAATTGATATCGGCTCT-3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pt2032_exp_rev</td>
<td>5’-CCTTACCCAGAGTAAAAGCC-3’</td>
</tr>
<tr>
<td></td>
<td>2171</td>
<td>Pt2171_exp_for</td>
<td>5’-AATACGTGGTGCGGTAGCTGG-3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pt2171_exp_rev</td>
<td>5’-GGTCIAAAAGAAGGATCCGAC-3’</td>
</tr>
<tr>
<td>PtNR</td>
<td>54983</td>
<td>PtNR_exp_for</td>
<td>5’-AATTGGAGTACTCGGCAAC-3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PtNR_exp_rev</td>
<td>5’-TTGTTAGGAAATGGTACGAC-3’</td>
</tr>
<tr>
<td>PtNAR1</td>
<td>13076</td>
<td>PtNAR1_exp_for</td>
<td>5’-TTGGTTTGCCCTTTGGTCTG-3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PtNAR1_exp_rev</td>
<td>5’-GCAATATTAGTGCCGCAAC-3’</td>
</tr>
</tbody>
</table>
qPCR amplification was performed using Fast SYBR Green Master mix with ROX (Applied Biosystems) in a final volume of 10 μL, using ViiA™ 7 Real-Time PCR System (Applied Biosystems). qPCR conditions used were as follows: 95 °C for 20 sec, 40 cycles at 95 °C for 1 sec and 60 °C for 20 sec, 95 °C for 15 sec, 60 °C 1 min, and a gradient from 60 °C to 95 °C for 15 min.

Data obtained were processed with the ViiA™ 7 Real-Time PCR system software. Fold-changes were obtained with the Relative Expression Software Tool-Multiple Condition Solver (REST-MCS) (Pfaffl et al., 2002) and values above ±2 were considered significative.

4.2.12. Vector for PtNPF1 heterologous expression in Xenopus laevis

*Xenopus laevis* oocytes are a commonly used heterologous system for the expression and functional characterisation of membrane proteins, such as ion channels or transporters, allowing to determine the specific substrate(s) transported by PtNPFs, their affinity for this substrate, their transport kinetics and specific conditions potentially acting on their function (Miller and Zhou, 2000).

The vector pGEM-Xho, a modified pGEM vector kindly provided by Dr. Lacombe (Institut des Sciences des Plantes de Montpellier - France), and containing the gene for ampicillin resistance, was used as backbone for cloning the full-length PtNPF1 sequence. PtNPF1 was amplified from genomic DNA, using primers PtNPF1_EcoRI_for 5’-ccgGAATTCATGACGACTCCCAGTGAGA-3’ and PtNPF2_XbaI_rev 5’-ctagTCTAGACTAGGAGGACGTTGCTCGCAA-3’, with underlined restriction site.

The full-length PtNPF1 sequence and vector pGEM-Xho were both cut with restriction enzymes EcoRI and XbaI in Buffer 3.1 (BioLabs), at 37°C for 2 hours, in order to generate sticky ends. The digested plasmid was then run on 1% agarose gel and the band corresponding to digestion was extracted using the GenElute™ Gel Extraction Kit (Sigma-Aldrich) and quantified with the Nanodrop. The PtNPF1 fragment was added to the plasmid backbone in a 3:1 ratio with an ideal backbone amount of 50 ng and incubated in the presence of 1 unit of T4 ligase enzyme (BioLabs) and company-supplied buffer.
(BioLabs) in a final volume of 20 μl, overnight at 16°C. The next day, tubes were incubated at 65°C for 10 minutes to deactivate ligase activity.

Transformations were carried out in the One Shot™ TOP10 Chemically Competent *E. coli* strain (Invitrogen) through heat shock: 100 μl of cells were added to 10 μl of ligase, put on ice for 10 minutes, exposed to shock at 42°C for 30 seconds, put on ice for 2 minutes. 1 ml of LB medium was added to cells and put for 1 hour at 37°C. Then, cells were plated on LB solid medium containing 100 μg/ml ampicillin (Sambrook et al., 1989) and put at 37°C overnight.

Diagnostic digestion was carried out on clones growing on selective medium and plasmid DNA was then extracted from positive clones, using GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich) and according to manufacturer’s protocol. Plasmidic DNA was sequenced, quantified using Qubit™ 2.0 Fluorometer (Life Technologies), and sent to Dr. Lacombe at the Institut des Sciences des Plantes de Montpellier (France) for the heterologous expression in *X. laevis* oocytes.

4.3. Results

4.3.1. A di-leucine-based motif important for vacuolar targeting found in PtNPF1

Previous bioinformatics analyses were carried out on PtNPF1 sequence, using the online tools LocTree3 and WolfPSort in order to predict its subcellular localisation, and using SignalP in order to detect possible signal targets (see paragraph 2.3.8 “Predicted subcellular localisation” in Chapter 2). From this extensive analysis, PtNPF1 was predicted to the plasma membrane and no signal target was identified. However, 25% of the diNPF sequences belonging to the same Clade II were predicted to the vacuolar membrane (Santin et al., 2021b). So further analyses were carried out to understand if PtNPF1 could show any signature of a possible localisation to the tonoplast.

For correct vacuolar targeting, important motifs are known from land plants and other eukaryotes. One of these is a conserved di-leucine-based motif, [D/E]xxxL[L/I], identified by Bonifacino and Traub (2003) and known to be essential for vacuolar targeting in plants such as *Arabidopsis* (Park et al., 2013) and in diatoms such as *P. tricornutum* (Schreiber et al., 2017). Through the online tool FIMO, a putative di-leucine
motif was detected close to the N-terminus of PtNPF1 sequence, in particular the residues [EHAPLL] from the 16th to the 21st aa (Fig. 4.2), with a p-value of $8.31 \cdot 10^{-5}$.

Figure 4.2. Phyre2 output showing the predicted secondary structure of the N-terminal region of PtNPF1 amino acid sequence. The di-leucine based motif representing the predicted vacuolar signal peptide is highlighted in red. The first predicted transmembrane helix (TMH 1) is indicated in gold.

4.3.2. Generation and selection of P. tricornutum PtNPF1 overexpressing strains

To examine the PtNPF1 cellular localisation, transgenic lines expressing the PtNPF1 protein fused to a YFP tag at the C-terminal under the regulation of the Lhcf2 promoter, or fused to a GFP tag at the N-terminal under the regulation of the PmH4 promoter, were generated (Fig. 4.3). To generate overexpressing strains, a co-transformation was performed, with two plasmids introduced together into the cells, one containing the gene conferring resistance to the antibiotic selection and the other one containing the full-length sequence of PtNPF1, in frame with a 3’ YFP or a 5’ GFP.

The choice of using fluorophores fused in one case at the C-terminal and in the other case at the N-terminal was determined by the possibility that the localisation signal could be hidden by the fluorescent tag. In fact, localization signals are often close to the N-terminus and can be masked or altered by fusing the protein with a tag: using plasmids with the fluorophore in a different position can help to avoid this issue and to eventually confirm obtained results of subcellular localisation. Generated plasmids for PtNPF1 overexpression are shown in Figure 4.3.
Figure 4.3. Vectors generated for PtNPF1 overexpression, through co-transformation with resistance plasmid: A) Lhcf2p-PtNPF1-YFP-Lhcf1t plasmid, with PtNPF1 upstream of YFP and controlled by the *P. tricornutum* Lhcf2 strong promoter; B) PmH4p-GFP-PtNPF1-Lhcf1t plasmid, with PtNPF1 downstream of GFP and controlled by the *P. multistriata* PmH4 constitutive promoter; C) Lhcf6p-Sh-Ble-Lhcf1t plasmid, with Sh-Ble gene conferring Phleomycin resistance (Falciatore et al., 1999).

As regards to *PtNPF1-YFP* overexpressing strains, after almost 20 days from biolistic transformation, colonies growing on selective medium were transferred in fresh medium and screened through PCR, using two pairs of primers to check the cassette insertion in *P. tricornutum* genome (Fig. 4.4B). Non-transformed wild-type strain was used as control of the selection and screening process. A total of 7 colonies contain the whole transgene insertion, out of 12 screened (around 60%) (Fig. 4.4C).
Figure 4.4. Selection and PCR screening of *PtNPF1-YFP* overexpressing strains. A) Positive C+ and negative C- control plates with non-transformed wild-type cells growing on F/2 medium without and with selection respectively, and four plates with transformed cells growing on selection. B) Schematic representation of the *PtNPF1-YFP* cassette with the two primer couples used, represented by red and green arrows and amplifying two partially overlapping regions. C) Agarose gels showing the results of the PCR amplification of fragments from promoter to the gene of interest *PtNPF1* and from *PtNPF1* to *YFP*. M represents the 1Kb marker, C+ the plasmid and C- the blank.

For *PtNPF1-YFP* overexpressing clones 1OE 4 and 1OE 9, which showed the whole cassette inserted, the *PtNPF1* gene expression levels were measured, in order to confirm the overexpression (Fig. 4.5). 1OE 4 shows a 5.12 ±0.27 -fold increase, while 1OE 9 shows a 4.21 ±0.21 -fold increase, compared to wild-type. These two clones were subsequently used for phenotypic analyses.
Figure 4.5. *PtNPF1* gene relative expression levels: A) agarose gel showing good quality RNA extracted from *P. tricornutum* wild-type and *PtNPF1* overexpressing strains 1OE 4 and 1OE 9; B) PCR performed with intron-spanning primers to check gDNA contamination (the amplicon from the gDNA containing intron is 150 bp, while the amplicon from the cDNA without intron is 90 bp). M represents the 1Kb marker. C) *PtNPF1* relative gene expression levels of overexpressing strains, normalized on the internal control *RPS* gene and compared to wild-type set as zero. Error bars represent the standard deviation of three technical replicates of two biological replicates.

Also, *GFP-PtNPF1* overexpressing strains were screened using two primer couples to check the whole cassette integration (Fig. 4.6A): among 10 positive colonies analysed through PCR, 4 show the whole transgene insertion, so 40% of the screened clones (Fig. 4.6B).
Figure 4.6. Selection and PCR screening of GFP-PtNPF1 overexpressing strains. A) Schematic representation of the GFP-PtNPF1 cassette with the two primer couples used, represented by red and green arrows and amplifying two partially overlapping regions. B) Agarose gels showing the results of the PCR amplification of fragments from promoter to GFP and from GFP to the gene of interest PtNPF1. M represents the 1Kb marker, C+ the plasmid and C- the blank.

4.3.3. PtNPF1 subcellular localisation

Subcellular localisation was determined by observing PtNPF1 overexpressing strains through fluorescence confocal microscopy. As shown in Figure 4.7, PtNPF1 is integrated into the tonoplast. This localisation pattern was observed in the two independent strains expressing PtNPF1 with C-terminal YFP, 1OE 4 and 1OE 9, and confirmed in other three PtNPF1-YFP independent strains which inserted the whole cassette from the first PCR screening.
Subcellular localisation of PtNPF1 in *P. tricornutum*. Confocal microscopy images of *P. tricornutum* cells expressing C-terminal YFP-tagged PtNPF1. First column: BF corresponds to Bright-Field; second column: chlorophyll *a* autofluorescence in red; third column: YFP fluorescence in green; fourth column: merged channels. Scale bar: 5 µm.

This localisation pattern was confirmed as well, in the strains expressing PtNPF1 with N-terminal GFP, in particular the two strains 1OE 29 and 1OE 31 (Fig.4.8).
Recently the dye MDY-64 was shown to mark the tonoplast of \textit{P. tricornutum} (Huang et al., 2016) and was used here to visualise this membrane (Huang et al., 2018). In Figure 4.9, stained wild-type cells show the same localisation pattern of the overexpressing strains, confirming the fluorescent fusion protein localisation on the tonoplast.
4.3.4. Vacuole behaviour in N starvation and different N sources

The vacuole is considered a storage organelle which allows diatoms to maintain their basal metabolism in different unfavourable conditions.

Morphological changes of the vacuoles in *P. tricornutum* wild-type grown respectively in normal and low NO$_3^-$ concentrations (882 and 50 µM) were analysed after 4 days using the MDY-64 tonoplast tracker (Figs. 4.10A-D). Cells own two or more vacuoles, the higher is their number the smaller is their size, covering around half of the total cellular area (Figs. 4.10A-D). The average number of vacuoles does not vary between the two different growth conditions, with value of 3.48 ± 1.74 in normal conditions and 3.69 ± 1.40 in N starvation (Fig. 4.10E). In fact, both conditions present cells with few big vacuoles (Figs. 4.10A and C) and cells with more small ones (Figs. 4.10B and D).

Similarly, the ratio between vacuole and total cell areas does not significantly change, with around 54% ± 9 of the cell volume occupied by vacuoles in normal conditions and around 58% ± 13 in N starvation (Fig. 4.10F). The number of cells observed for 172
morphological parameters per each condition is reported in the Supplementary Table S4.2. However, intracellular NO$_3^-$ content, which is mainly stored in the vacuole, shows a significant reduction from 1.23 ± 0.04 pg/cell in normal conditions to 0.42 ± 0.07 pg/cell in N starvation condition, consistent with the cellular consumption (Fig. 4.10G).

**Figure 4.10.** Morphological parameters and intracellular NO$_3^-$ content of *P. tricornutum* wild-type in different N concentrations. A-D) Confocal microscopy images of *P. tricornutum* wild-type cells stained with the vacuole tracker MDY-64, after 4 days in normal N conditions (A and B) and N starvation (C and D). Chlorophyll a in red, tonoplast tracker MDY-64 in cyan. Scale bar: 5 µm. E) number of vacuoles and F) ratio between total vacuole area and total cellular area. The values are shown as mean ± SD (N= see Supplementary Table S4.2). G) intracellular NO$_3^-$ content in *P. tricornutum* cells grown in normal N condition and N starvation. The values are shown as mean ± SD (N=3 biological replicates analysed for intracellular NO$_3^-$ content per each condition). * indicates $p < 0.05$.

Morphological features of the vacuoles were investigated in *P. tricornutum Ptnpf1* knock-out mutants (see paragraph 3.3.5 “Ptapt-Ptnpf1 knock-out mutants generation and screening” in Chapter 3) and *PtNPF1-YFP* overexpressing strains as well, in normal conditions (882 µM NO$_3^-$) and after 2, 4 and 7 days of N starvation, using the MDY-64
tonoplast tracker, or YFP fluorescence for overexpressing strains (Fig. 4.11). The number of cells observed for morphological parameters per each strain and per each condition is reported in Supplementary Table S4.2.

Figure 4.11. Confocal microscopy images of *P. tricornutum* wild-type, *Ptnpf1* knock-out mutant 2.9 and *PtNPF1*-YFP overexpressing strain 1OE 4 after 0, 2, 4 and 7 days of N starvation. The number of total observations is reported in the Supplementary Table S4.2. Chlorophyll *a* is shown in red, tonoplast tracker MDY-64 in cyan. For *PtNPF1* overexpressing strain 1OE 4, tonoplast visualisation was performed through the fluorescence of the YFP tag, in green. Scale bar: 5 µm.

No differences have been observed in vacuoles number or size, neither between strains nor during N starvation time (Figs. 4.12A-B), similarly to what previously observed in wild-type (Figs. 4.10E-F). The chloroplast area of all the strains was measured to confirm the N starvation condition. Results show a progressive reduction of the plastid volume in all the strains, with a maximum value of 54.8% ± 8.7 in the wild-type and a minimum value of 46.3% ± 9 in the 1OE 4 strain at the beginning of the experiment, and a maximum value of 11.2% ± 2.4 in the wild-type and a minimum value of 9.2% ± 3 in the 1KO 2.9 strain after 7 days of N starvation (Fig. 4.12C).
Then, the intracellular NO$_3^-$ content of *P. tricornutum* wild-type, 1KO 2.9 and 1OE 4 strains was measured in a different N source, 882 µM NH$_4^+$, to evaluate the preference of *P. tricornutum* for N source utilisation. In fact, it has been proposed that diatoms could be NO$_3^-$ specialist (Glibert et al., 2016; Lomas and Glibert, 2000), as previously described in paragraph 1.3.4 “Nitrogen transporters in diatoms” in Chapter 1.

During two parallel growth curves, in NO$_3^-$ and NH$_4^+$ respectively, the intracellular NO$_3^-$ does not change neither between the two conditions nor between different strains, suggesting a preference in taking up different N sources from external environment when it is present, rather than using internal stored resources (Fig. 4.13).
Intracellular NO$_3^-$ content in *P. tricornutum* wild-type, *Pt*nsfl knock-out mutant 1KO 2.9 and *PtNPF1*-YFP overexpressing strain 1OE 4 cells grown in 882 µM NaNO$_3$ and 882 µM NH$_4$Cl as alternative N sources. The values are shown as mean ± SD (N=3 biological replicates per each strain and condition).

4.3.5. *Pt*nsfl knock-out phenotype in response to N repletion

To investigate the impact of *PtNPF1* overexpression and knock-out in the physiology of *P. tricornutum*, different experiments with wild-type, *PtNPF1*-YFP overexpressing strains 1OE 4 and 9 and *Pt*nsfl mutants 1KO 2.8 and 2.9 (see paragraph 3.3.5 “Ptapt-*Pt*nsfl knock-out mutants generation and screening” in Chapter 3) were performed.

First of all, *PtNPF1* was classified as NO$_3^-$ transporter but no differential expression of the *PtNPF1* gene was observed in different NO$_3^-$ concentrations or different N sources, suggesting environmental N does not influence *PtNPF1* gene expression (see paragraph 2.3.5 “Expression patterns of NPFs in *P. tricornutum*” in Chapter 2). To confirm these data, different *P. tricornutum* strains were grown in two different NO$_3^-$ concentrations, 882 µM set as normal condition and 50 µM set as N starvation condition: no difference in growth has been observed in different strains (Fig. 4.14C-D). Then, different N sources were tested, in particular the same concentration of NO$_3^-$, set as reference condition, NH$_4^+$ and urea. Generally, cells were shown to grow slightly faster in NH$_4^+$ than NO$_3^-$ and urea, as more cost-effective from an energy point of view; again, no differences are present between different strains (Fig. 4.14E-H).
Figure 4.14. Growth curves of *P. tricornutum* wild-type, *PtNPF1-YFP* overexpressing strains and *Ptnpf1* mutants in different NO$_3^-$ concentrations and N sources: A-B) 882 µM NaNO$_3$, C-D) 50 µM NO$_3$, E-F) 882 µM NH$_4^+$, G-H) 882 µM urea.
50 µM NaNO₃, E-F) 882 µM NH₄Cl and G-H) 882 µM urea. *In vivo* chlorophyll a fluorescence was used for measurements as proxy for growth, and natural logarithm of growth curves was calculated. A knock-out mutant for a different gene (2KO 1.15), generated with the same proteolistic protocol, was used as additional control strain. In all the graphs, error bars represent the standard deviation of three biological replicates. a.u. arbitrary units. * indicates *p* < 0.05; ** for *p* < 0.01 and *** for *p* < 0.001.

Then, a time-course experiment was performed, in which both wild-type, *PtNPF1-YFP* overexpressing strains and *Ptnpf1* knock-out cells were initially grown on 882 µM NaNO₃, transferred to 50 µM NaNO₃ media for 4 days, subsequently repleted in 882 µM NaNO₃, and periodically sampled. During the first four days of N starvation, no growth difference has been observed between different strains, as previously seen (Figs. 4.14C and D). However, after NO₃⁻ repletion, wild-type and *PtNPF1-YFP* overexpressing strains start growing, reaching fast the exponential phase, while both *Ptnpf1* clones 2.8 and 2.9 show a growth delay of about one day (Fig. 4.15A-B).

![Figure 4.15](image)

**Figure 4.15.** Growth curves of *P. tricornutum* wild-type, *PtNPF1-YFP* overexpressing strains and *Ptnpf1* mutants during N starvation and repletion experiment: A) *In vivo* chlorophyll a fluorescence was used for measurements as proxy for growth; B) natural logarithm of the *in vivo* chlorophyll a fluorescence; C) liquid cultures used for the experiment, photographed two days after N repletion (on day 6). Labels indicate the different strains. Note the paler colour of flasks of the *Ptnpf1* knock-out mutants. A knock-out mutant for a different gene (2KO 1.15), generated with the same proteolistic protocol, was used as additional control strain. In all the graphs, error
bars represent the standard deviation of three biological replicates. a.u. arbitrary units. * indicates $p < 0.05$; ** for $p < 0.01$ and *** for $p < 0.001$.

To test if the delay in growth is due to the N or to a general nutrient stress, a P starvation and repletion experiment was performed, growing cells with no P for four days and then resupplying them with 36 µM NaH$_2$PO$_4$·H$_2$O. No growth difference has been observed between strains (Fig. 4.16A), suggesting a N specific response. The same experiment was repeated using other N sources such as NH$_4^+$ or urea, resulting in the same growth delay of Ptnpf1 mutants after N repletion (Fig. 4.16B-D), and suggesting the N response observed could be not environmental NO$_3^-$ specific.

Figure 4.16. Growth curves of *P. tricornutum* wild-type, PtNPF1-YFP overexpressing strains and Ptnpf1 mutants during P or different N sources starvation and repletion experiments: A) PO$_4^{3-}$ starvation and repletion, B) NO$_3^-$ starvation and repletion, and C) NH$_4^+$ starvation and repletion. *In vivo* chlorophyll $a$ fluorescence was used for measurements as proxy for growth. A knock-out mutant for a different gene (2KO 1.15), generated with the same proteolistic protocol, was used as additional control strain. In all the graphs, error bars represent the standard deviation of three biological replicates. a.u. arbitrary units. * indicates $p < 0.05$; ** for $p < 0.01$ and *** for $p < 0.001$. 

179
To explain the growth delay of *Ptnpf1* knock-out mutants, extra- and intracellular NO$_3^-$ content was measured. Results shown in Figures 4.17C and 4.17D highlight a different uptake of medium NO$_3^-$ by different strains. In particular, after NO$_3^-$ repletion wild-type starts consuming external NO$_3^-$, whose concentration decreases in the external medium and increases in the intracellular environment. On the other hand, *Ptnpf1* mutants are not immediately able to use extracellular NO$_3^-$ provided, and cells need more time to internalise it and increase intracellular NO$_3^-$ concentration (Figs. 4.17C-D). Even if the difference in extracellular NO$_3^-$ between wild-type and *Ptnpf1* mutants could be a direct consequence of the delay in N uptake or an indirect effect due to the impaired concentration of cells which effectively uptake N from the environment, intracellular NO$_3^-$ content is normalised on cell concentration, suggesting this imbalance could be mainly due to an impaired N uptake.

**Figure 4.17.** Growth and NO$_3^-$ concentrations in wild-type and *Ptnpf1* knock-out cells. A) Cell counts for wild type, *PtNPF1-YFP* overexpressing strains and *Ptnpf1* knock-out strains. A knock-out mutant for a different gene (2KO 1.15), generated with the same proteolistic protocol, was used as additional control strain. B) Measurement of *in vivo* chlorophyll a autofluorescence providing a proxy for growth. C) Extracellular NO$_3^-$ content measured by UV spectrophotometry. D) Intracellular NO$_3^-$ extracted from cells and measured by UV spectrophotometry. In all the graphs, error bars represent the standard deviation of three biological replicates. a.u. arbitrary units. * indicates $p < 0.05$; ** for $p < 0.01$ and *** for $p < 0.001$. 

180
Another interesting point is that the phenotype observed is transient: after one to two days from NO$_3^-$ repletion, *Ptnpf1* knock-out strains recover the phenotype and start to grow with the same rates of the wild-type. There is indeed a delay in growth and delay in NO$_3^-$ uptake, rather than a growth reduction. The transient phenotype suggests a potential redundancy with other key players that could be part of the molecular mechanism allowing cells to recover from starvation once N is available again in the medium.

### 4.3.6. Gene expression analyses

During the N starvation and repletion experiment, the expression levels of genes involved in N metabolism and transport were analysed. In particular, cells of *P. tricornutum* wild-type, *Ptnpf1* knock-out strain 1KO 2.9, and *Ptnpf2* knock-out strain 2KO 1.15 used as additional control, were collected at the beginning of the growth curve (T0), after 4 days of N starvation (T4), the day after NO$_3^-$ repletion when the growth difference was visible (T5) and after three days from N repletion (T7).

The relative expression level of *PtNPF1* does not change between wild-type and 2KO 1.15 during time, as well as *PtNPF2* expression levels between wild-type and 1KO 2.9, suggesting as expected that there is no redundancy between *PtNPF* genes (Fig. 4.18A and B).

Then the expression levels of *P. tricornutum* high-affinity NO$_3^-$ transporters (*PtNRT2s*) were investigated. Among six *PtNRT2s*, three were predicted to localise on the tonoplast, while the others to the plasma membrane, even if not still experimentally validated (Busseni et al., 2019). Moreover, from previous literature data, the putative tonoplast-localised *Pt54560* and the putative plasma membrane-localised *Pt26029* and *Pt54101* were shown to be upregulated in N starvation condition (Levitan et al., 2015).

Three out of six *PtNRT2s* show a slower response after N repletion in *Ptnpf1* knock-out mutants compared to wild-type.

In detail, the expression levels of *PtNRT2s* predicted to the tonoplast, namely *Pt54560*, *Pt40691* and *Pt2171* were analysed: *Pt54560*, upregulated in N starvation at T4, shows a fast decrease in the fold-change expression in wild-type after N repletion at T5, while in the *Ptnpf1* mutant this decrease is slower, maintaining high expression levels also in T5 (Fig. 4.18C). On the other hand, *Pt40691* and *Pt2171* do not show any difference in gene expression levels over time and among different strains (Figs. 4.18D-E). Then, the expression levels of *PtNRT2s* predicted to the plasma membrane, namely *Pt26029*,
Pt54101 and Pt2032 were analysed as well. Pt26029 and Pt54101 levels, upregulated in all strains at T4, show a fast decrease in wild-type and a slower decrease in Ptnpf1 mutants after N repletion, presenting a significant difference compared to wild-type at T5 (Figs. 4.18F-G). Pt2032 does not show any difference in fold-change expression over time and between strains (Fig. 4.18H).

Moreover, genes involved in N assimilation were tested, such as the Nitrate Reductase (PtNR) which catalyses the reduction of NO$_3^-$ into NO$_2^-$, and the Nitrite Transporter 1 (PtNAR1) which transports NO$_2^-$ into the chloroplast for further reduction steps: both genes show an upregulation in response to N starvation at T4, followed by a decrease in their expression levels after N repletion in T5, with no differences between wild-type and the Ptnpf1 strain (Figs. 4.18I-L).

Finally, the gene encoding for the Chloride Channel 3 (PtCLC3), which is a putative vacuolar chloride channel identified from previous transcriptomic studies on N starvation (Scarsini et al., 2022), was tested, in order to detect any other player involved in N reallocation from the vacuole. However, no differences have been observed between strains (Fig. 4.18M).
Figure 4.18. Relative expression levels of genes involved in N transport and metabolism in *P. tricornutum* wild-type, *Ptnpf1* knock-out mutant 1KO 2.9 and *Ptnpf2* knock-out mutant 2KO 1.15 strains during N starvation and repletion experiment. A-B) *PtNPFs*, C-D-E) *PtNRT2s* predicted to the tonoplast, F-G-H) *PtNRT2s* predicted to the plasma membrane, I-L) *PtNR* and *PtNAR1* genes involved in N assimilation, M) *PtCLC3* encoding a tonoplast chloride channel. T0 represents the beginning of the experiment, before the N starvation beginning, T4 indicates N starvation reached after four days in 50 µM NaNO₃, T5 indicates the day after repletion with 882 µM NaNO₃ (5 days from the beginning of the experiment) and T7 indicates three days after N repletion (7 days from the beginning of the experiment). Wild-type at T0 was set as reference condition at zero. Error bars represent the standard deviation of three technical replicates of two biological replicates. Differences ≥ 2 fold change, indicated by the horizontal dashed lines, are considered statistically significative.
4.4. Discussion

Previous studies have demonstrated that diatoms undergo dramatic metabolic changes in response to fluctuations in N availability (Alipanah et al., 2015). To cope with these variations, diatoms have evolved the capability to mobilise N and its organic and inorganic compounds, such as NO$_3^-$, NO$_2^-$ and NH$_4^+$, from the environment and deliver them to the different compartments of the cell. However, the cellular mechanisms involved in N mobilisation and intracellular distribution in diatoms are still not well understood. Recently, several studies have investigated whole cell transcriptomic responses to N deprivation (Alipanah et al., 2015, 2018; Scarsini et al., 2022), intracellular N fluxes between organelles (Smith et al., 2019) and role of some key enzymes of N metabolism (McCarthy et al., 2017). Here, new insights on vacuole behaviour were provided, and the tonoplast localised PtNPF1 transporter was characterised as putatively involved in molecular mechanisms regulating N storage and reallocation.

N starvation represents one of the main stressing conditions for diatoms. In this condition, N previously accumulated in the vacuole mainly in the form of NO$_3^-$, becomes newly bioavailable for the cell (Bender et al., 2014; Stief et al., 2013; Lomas and Glibert, 2000). In N starvation, $P$. tricornutum vacuoles do not change in morphology, neither in number or in size, differently to what was observed in phosphate deprivation (Dell’Aquila et al., 2020). However, the intracellular NO$_3^-$ content, which is mainly stored in the vacuole, dramatically decreases in N starvation, confirming a reallocation of resources and their assimilation and consumption. Since morphology is not affected, it is possible that the changes in intracellular NO$_3^-$ correspond to a different texture or biochemical composition of the vacuole (McCarthy et al., 2017; Raven and Beardall, 2022). In fact, the vacuole represents approximately half of the cell volume, and it also plays a critical role in osmotic regulation, predator defence and buoyancy control (Behrenfeld et al., 2021). Changes in vacuole size and number have to take into account all these different needs, and a chemical modification rather than a size change can be the consequence of an established self-regulation (Raven and Beardall, 2022).

Another interesting observation is about the preference for different N sources available. Different results and hypotheses on phytoplankton preference for NH$_4^+$ or NO$_3^-$ have long been proposed (Glibert et al., 2016). Results here described (Fig. 4.13) suggest that, when NH$_4^+$ rather than NO$_3^-$ is given from the external environment, diatoms prefer to take it
from the medium while maintaining intracellular N stored, so that intracellular NO$_3^-$ does not vary in this condition. This confirms the central tenet of the relationships and interactions between NH$_4^+$ and NO$_3^-$, which considers NH$_4^+$ to be the energetically preferred form of N for phytoplankton uptake, as it is in the correct oxidation state for subsequent incorporation into nucleic acids, amino acids and proteins, while NO$_3^-$ needs to be further reduced (Dortch, 1982; Dortch et al., 1985).

Even if, among various inorganic N species, NH$_4^+$ is preferably taken up and directly consumed by phytoplankton (Mikaelyan et al., 2015), several considerations need to be taken into account. First of all, it has also been documented that under conditions of highly elevated NH$_4^+$ concentrations, typically exceeding several tens to hundreds of μM, both the total N taken up and the overall growth can be suppressed rather than enhanced, because of the redox imbalances and a surplus of reductant when NH$_4^+$ is in excess (Glibert et al., 2016). Moreover, the most bioavailable N source in the ocean is represented by NO$_3^-$. This, together with the strong capability of diatoms to store N in vacuole, mainly as NO$_3^-$ (Behrenfeld et al., 2021; Kamp et al., 2011), which needs to be taken up from the external environment during favourable periods, indicates an equally important role of NO$_3^-$ in diatom N metabolism. In fact, diatoms can take strong advantage from variable NO$_3^-$ availability, being one of the most competitive phytoplanktonic groups forming blooms in upwelling and coastal areas (Dutkiewicz et al., 2015a; Tréguer et al., 2018). Collectively, the observations about preferential use of N sources, and about repression and/or toxicity at high concentrations, have led to the hypothesis that an equilibrium can be reached through co-provision of both NH$_4^+$ and NO$_3^-$ (Glibert et al., 2016), whose fine regulation can induce synergistic growth but which still needs to be clearly understood.

To sense and respond to N changings, diatoms own a complex sensing and transport system which allows them to cope with different situations (Rogato et al., 2015). In particular, as the most common forms of inorganic N available for phytoplanktonic organisms are NO$_3^-$ and NH$_4^+$, NO$_3^-$ transporters (NRT2s and NPFs) and NH$_4^+$ transporters (AMT) are the most interesting player involved in N exchanges (Rogato et al., 2015). However, most of these mechanisms are still little known as well as proteins and transporters involved have not been characterised yet.

Diatom AMTs were all predicted to be plasma membrane located (Busseni et al., 2019), a result consistent with the lack of reported storage mechanisms for NH$_4^+$ in the vacuole.
of diatoms (McCarthy et al., 2017). On the other hand, diatoms have a set of high- and low-affinity NO$_3^-$ transporters, predicted to be distributed not only on the plasma membrane but also on other cellular membranes including tonoplast (Busseri et al., 2019; Santin et al., 2021b). Although they could be crucial actors for intracellular reallocation of N, no functional characterisation has been done so far for these transporters, and their role still needs to be defined. Using _P. tricornutum_ as model diatom for this study, one of the two low-affinity NO$_3^-$ transporters, PtNPF1, has been further investigated. It belongs to Clade II diNPFs, evolutionary and structurally close to higher plant NPFs (see paragraph 2.3.2 “Phylogeny of diNPFs” in Chapter 2).

Previous _in silico_ prediction of diNPFs subcellular localisation, allowed to observe an interestingly high number of diNPF sequences predicted to the tonoplast, all belonging to the Clade II (see paragraph 2.3.8 “Predicted subcellular localisation” in Chapter 2) (Santin et al., 2021b). From results described in this Chapter, a vacuole targeting motif has been recognised in _P. tricornutum_ PtNPF1 amino acid sequence and, through overexpression of PtNPF1 fused with fluorescent tags, PtNPF1 localisation on tonoplast has been experimentally confirmed.

The hypothesis is that PtNPF1 can transport NO$_3^-$ from the vacuole towards cytosol, suggesting that the transporters could be responsible for NO$_3^-$ reallocation inside the cell. This hypothesis is based on the PtNPF1 homology with low-affinity plant NPFs, active at high NO$_3^-$ concentrations, and the high NO$_3^-$ concentrations which can be reached in diatom vacuole, up to 60 mM (Kamp et al., 2011). Moreover, considering structural homology with plant AtNPF6.3 (see paragraph 2.3.7 “Structural modelling of diNPFs” in Chapter 2), PtNPF1 would be supposed to fold and insert in the membrane, in this case the tonoplast, with the two ends located in the cytosol, so transporting the substrate from the vacuole to the cytosol, and not in the other direction, for which other two transmembrane helices should be requested to overturn the protein (Longo et al., 2018; Yan et al., 2013).

PtNPF1 involvement in vacuolar NO$_3^-$ efflux towards cytosol could be of great importance to N reallocation and efficient utilisation by _P. tricornutum_ cell, explaining how N taken up and stored in vacuoles can be subsequently mobilised. Following this hypothesis, PtNPF1 could mediate NO$_3^-$ efflux similarly to AtNPF5.11, AtNPF5.12 and AtNPF5.16 which are responsible for vacuolar NO$_3^-$ release in _Arabidopsis_ (He et al., 2017).
Since PtNPF1 is not differentially expressed in different concentrations of NO$_3^-$ (see paragraph 2.3.5 “Expression patterns of NPFs in *P. tricornutum*” in Chapter 2), and the *Ptnpf1* mutants do not show a different growth phenotype compared to the wild-type in N starvation, a redundancy can be present, with other transporters directly involved in the reallocation of intracellular NO$_3^-$ during unfavourable periods.

However, the growth delay together with the delay in N uptake observed in *Ptnpf1* knock-out mutants after N repletion means that PtNPF1 plays an important role in N metabolism and transport. In detail, extracellular NO$_3^-$ remains higher in *Ptnpf1* knock-out mutants compared to wild-type in the first days after N repletion. This could be an indirect effect due to the lower cell concentration in *Ptnpf1* mutants, which consequently take up less NO$_3^-$ available in the medium, but more likely it could be a direct effect of reduced NO$_3^-$ uptake. The second option is supported by intracellular NO$_3^-$ content, normalised per cell, and by a reduction of plasma membrane *PtNRT2s* levels of induction (see *Pt26029* and *Pt54101* expression levels in T5 in the previous paragraph 4.3.6 “Gene expression analyses”).

So, the growth and the N uptake delay in *Ptnpf1* knock-out mutants suggests a role of PtNPF1 in signalling between different cell compartments (Fig. 4.19). NO$_3^-$ is normally taken up and accumulated in the vacuole, to be released in the cytosol and assimilated during unfavourable periods or conditions (Kamp et al., 2015), such as N starvation. When NO$_3^-$ is present again in the environment, the NO$_3^-$ flux inside the cell needs to be inverted, blocking vacuolar NO$_3^-$ release and newly allowing the NO$_3^-$ uptake from the external environment. This implies that there is a communication between the plasma membrane and the tonoplast, with key proteins that act as sensors and transfer the signal between different membranes. These signalling pathways trigger responses that result in the NO$_3^-$ flux inversion (Fig. 4.19), depending both on the physiological conditions of the cell, such as the concentration of NO$_3^-$ accumulated into the vacuole or directly available in the cytosol, and on the surrounding medium, such as NO$_3^-$ concentration, other N compounds and/or light (Kamp et al., 2011). The observed growth delay after repletion with N sources different from NO$_3^-$ supports the idea that this is a N-related response but not NO$_3^-$-specific, so probably not (or not only) depending on environmental N sources.
Sensing N availability in different extra- and intracellular compartments requires signalling pathways for communication and response. Published literature on NO$_3^-$ signalling mechanisms can be divided into three categories: i) transcription factors controlling NO$_3^-$ responsive gene expression (Smith et al., 2019), ii) kinase phosphorylation of NO$_3^-$ responsive proteins (Tan et al., 2020), and iii) NO$_3^-$ transporters acting as both sensor and transporter in higher plants (Gojon et al., 2011). In particular, interesting dual sensor/regulator mechanisms have been observed in bacteria NarX and NarL (Mangalea and Borlee, 2022) and in higher plant NPF6.3 (Gojon et al., 2011), where a sophisticated mechanism allows the phosphorylation of the Thr101 residue by serine/threonine kinases and regulates affinity for NO$_3^-$ (Liu and Tsay, 2003). Thr101 has been shown to be conserved in many, but not all, of the diNPF sequences (Santin et al., 2021b). In particular, in _P. tricornutum_ NPF1, it is replaced by a serine (Ser260) (see paragraph 2.3.7 “Structural modelling of diNPFs” in Chapter 2), which could allow signal transduction by kinases similarly to Thr101 in _Arabidopsis_ NPF6.3. Nevertheless,
signalling pathways and ion fluxes are still poorly understood in diatoms (Smith et al., 2019), and to reconstruct these processes is one of the main challenges, considering that PtNPF1 could play a key role in this context.

The delay in growth and N uptake of \textit{Ptnpf1} mutants compared to wild-type cells after N repletion can be explained through the interruption of the signalling pathway that occurs between the plasma membrane and the vacuole, in the first moments after environmental NO$_3^-$ concentration increases. This is supported by the upregulation of three out of six \textit{PtNRT2s} in \textit{Ptnpf1} mutants not only during N starvation but also in the growth delay period just after N repletion, suggesting that N transport regulation becomes slower in the knock-out. In particular, Pt54560 is a PtNRT2 predicted to the tonoplast, its upregulation during N starvation suggests that it can be able to transport NO$_3^-$ out of the vacuole, and its relative expression level after N repletion, lower than N starvation but still high compared with wild-type in T5, suggests that the cell still prefers to reallocate vacuolar NO$_3^-$ rather than taking up the extracellular one.

The slower regulation of \textit{PtNRT2s} in \textit{Ptnpf1} mutants compared to wild-type, parallel to the delay in NO$_3^-$ accumulation after repletion, could also have another explanation. There could be a common mechanism controlling both NO$_3^-$ influx and \textit{PtNRT2} genes expression, similarly to higher plants NRT2s (Fraisier et al., 2000; Pii et al., 2016). They have a two-levels control: a post-translational control mainly based on the accessory protein NAR2 protein synthesis (Pii et al., 2016), and post-transcriptional events affecting abundance and/or activity of the NRT2 proteins, such as the regulation of transcript stability (Laugier et al., 2012). Since diatom genomes do not encode for NAR2 (Rogato et al. 2015), it is possible that PtNRT2s are under a post-transcriptional control, which could play a role in their response to environmental cues and could explain the similar trend in responses in transcript levels of \textit{PtNRT2} genes and NO$_3^-$ influx.

An unbalanced regulation of NO$_3^-$ transport system, but not a different regulation of N assimilation pathway, has been observed in \textit{Ptnpf1} mutants after N repletion, suggesting that different regulators can play a role in controlling N fluxes and metabolism inside the cell (Fig. 4.19). This is still well known for higher plants, with many components, such as transporters, receptors, second messengers and transcription factors, localised in different cellular compartments or free to move between them, involved in mediating short- and long-term metabolic responses to N (Lamig et al., 2022).
Moreover, despite the delay in growth and N uptake as well as a slower regulation of two PtNRT2 expression levels, after 4 days from N repletion all the strains reached the same physiological condition, indicating the phenotype observed was transient and other mechanisms occurred to shape diatom metabolism. A similar response was already observed by Tan (2020) who, in the first functional characterisation of the *P. tricornutum* high-affinity transporter Pt26029, observed a photosynthetic efficiency delay in knock-out response, with a subsequent recovery of the phenotype. This is probably due to redundancy of key players regulating N transport and metabolism in diatoms. From results described in this Chapter, redundancy of NO$_3^-$ transporters active at different substrate concentrations and localised in different compartments, could allow the restoration of NO$_3^-$ transport, albeit with a delay, but confers a crucial role to PtNPF1 in the NO$_3^-$ flux inversion in response to changing conditions.

Understanding the signalling pathways that transduce intra- and extra-cellular N signals into physiological responses is an active field of research, and can help to explain the reason of diatom success in the fluctuating and changing oceanic environment. Here, we propose PtNPF1 as a possible regulator of N intracellular dynamics, involved in controlling N fluxes between storage organelles and active assimilation compartments. This is an important brick that can help reconstruct the complex N metabolism in diatoms. However, many signals such as the recovery of the phenotype indicate that many key players have yet to be identified. In fact, identifying components of N signalling and understanding how they operate in different environmental and physiological conditions is critical to shed light on the ecological basis of diatom success.
Chapter 5: *P. tricornutum* NPF2 functional characterisation

Photosynthetic measurements and bicarbonate assay were performed in Dr. Angela Falciatore laboratory, in collaboration with Dany Croteau, Dr. Erik Jensen and Dr. Benjamin Bailleul (Institut de Biologie Physico-Chimique, IBPC, UMR 7141 - Chloroplast Biology and Light-sensing in Microalgae Laboratory at CNRS - Sorbonne Université in Paris, FR), thanks to the *EMBO Scientific Exchange Grant 9344*, on the research project "Exploring the role of diatom low affinity nitrate transporters (diNPFs) in photosynthesis and intracellular nutrient homeostasis."
**Abstract**

Cellular ion and pH homeostasis are fundamental for metabolic reactions, protein stability, signalling and transport mechanisms, which all require specific pH conditions. To control and balance intracellular ion concentrations, organisms need to constantly sense and respond to both extracellular environment and cell compartments requests. This finely tuned system has been deeply studied in many organisms such as higher plants and animals, but for diatoms, living in a highly variable environment, the key regulators of intracellular homeostasis have not been identified and characterised yet.

In particular, diatoms own a four-membrane chloroplast, derived from a secondary endosymbiosis event, where ion and pH regulation play a crucial role in controlling photosynthetic and photoacclimation processes together with carbon fixation. Molecular mechanisms allowing ions to move across these membranes can shed light on many aspects of diatom physiology.

Here, *P. tricornutum* Low-affinity Nitrate Transporter NPF2 is proposed to localise in a specific compartment between chloroplast membranes and to putatively play a role in proton transport across those membranes and intracellular pH regulation.
5.1. Introduction

Cellular homeostasis and regulation of ions concentration and pH are fundamental for the viability of all eukaryotic cells. In fact, metabolism, protein stability, ion channel activity, compartmental integrity, signalling and membrane trafficking events exhibit stringent pH requirements (Shen et al., 2013). Intracellular pH is strictly connected with the external environment but also with internal cell physiology. It varies within different intracellular compartments and generates proton (H\(^+\)) gradients which act as the fuel necessary to regulate specific processes and general metabolism of the cell, guaranteeing its viability (Shen et al., 2013; Feng et al., 2020). For example in higher plants, cytosolic pH measurements generally report values of approximately 7.2, with transient excursions of up to 0.3 pH units associated with signalling or photosynthesis onset in embryophytes (Taylor et al., 2012; Feng et al., 2020; Roos and Boron, 1981).

Only few studies have been carried out on intracellular pH on diatoms, unicellular photosynthetic organisms representing one of the major and most successful phytoplanktonic groups living in the ocean. In detail, \textit{P. tricornutum} reported an average intracellular pH of 7.6 at an extracellular pH of 7.5, even if the method used in this study did not permit determination of the pH of specific compartments (Burns and Beardall, 1987). Other methods allowed to obtain only relative intracellular pH changes in diatoms (Shi et al., 2019; Zhang et al., 2016). Cytoplasmic pH in another diatom, \textit{Thalassiosira weissflogii}, was shown to vary between 7.2 and 7.5 over the range of oceanic pH of 7.8–8.4 (Hervé et al., 2012) and to be generally similar to the cytoplasmic pH of other algae (Hopkinson, 2014).

Intracellular pH regulation involves several other ions with a buffering capacity, like phosphate (PO\(_4^{3-}\)) and hydrogenocarbonates such as bicarbonate (HCO\(_3^-\)), besides H\(^+\), and proteins playing an important role in controlling cytosolic pH. Moreover, intracellular pH regulation occurs through metabolic ions consumption and production together with transmembrane ion transport and compartmentalisation (Taylor et al., 2012).

For example, in animal cells pH homeostasis is largely mediated through a combination of HCO\(_3^-\) buffering activity and Na\(^+\)-coupled co-transporters (e.g., Na\(^+\)-H\(^+\) exchangers). On the other hand, higher plants and many freshwater algae use a “biophysical pH-stat” mainly characterised by H\(^+\) exchanges, involving plasma membrane P-type H\(^+\) ATPases, vacuolar V-type ATPases and/or pyrophosphatases (V-PPases) which translocate H\(^+\) to
the external or vacuolar medium respectively, so regulating intracellular pH. In tissues and/or organelles that do not have access to the external sink, a “biochemical pH-stat” may operate, that means a metabolic intracellular pH regulation by biosynthesis and degradation of organic acids (Taylor et al., 2012).

For marine phytoplankton, the mechanisms which regulate intracellular pH are still poorly understood, but they seem to employ an animal-like uptake system, together with a plant-like buffering activity. The uptake of nutrients, including inorganic forms of nitrogen (nitrate, NO\textsubscript{3}\textsuperscript{−}, or ammonium, NH\textsubscript{4}\textsuperscript{+}) and carbon (such as HCO\textsubscript{3}\textsuperscript{−}), is often coupled with the transport of ions such as Na\textsuperscript{+}, H\textsuperscript{+} or Cl\textsuperscript{−}, causing primary effects on ion and pH balance in cells (Taylor et al., 2012). Furthermore, metabolic processes resulting in the generation and consumption of H\textsuperscript{+} in marine phytoplankton include the production of large quantities of dimethylsulphonioiopropionate (DMSP) from sulphate, the uptake and use of HCO\textsubscript{3}\textsuperscript{−} in the Carbon Concentrating Mechanisms (CCMs), and the generation of intracellular acidic compartments for silicification in diatoms, which all influence the balance of H\textsuperscript{+} fluxes between cellular compartments (Taylor et al., 2012).

A special intracellular compartment where pH plays a crucial role is the chloroplast. The difference in pH between the plastidial stroma and the thylakoid lumen creates a transmembrane pH gradient or proton potential (ΔpH), that is at the same time one of the two components of the electrochemical proton gradient, or proton motive force (pmf), which drives chloroplast ATP synthase for ATP synthesis, but also a master regulator of photosynthesis. The lumenal pH has been estimated at 5.8–6.5 under normal light conditions and 4.5–4.8 under high light conditions (Kramer et al., 1999) while stromal pH in the dark was reported to be around 7, increasing to 7.8–8.0 in the light (Trinh and Masuda, 2022). So, the ΔpH across thylakoid membranes can show variations, reaching a maximum ΔpH of 1.8–2.1 under steady light (Tikhonov et al., 2008).

At the same time, pH is a master regulator of photosynthesis (Lepetit et al., 2022). In response to light intensity and CO\textsubscript{2} concentration in the environment, chloroplasts can perform photosynthesis effectively and finely tune the mechanisms that protect against unfavourable conditions (Trinh and Masuda, 2022). In this context pH acts as a signal or messenger which reflects changes in the external and intracellular environment to control protective mechanisms against photodamage. As example, the lumenal pH acts as photosynthetic control, as the electron transport rate through cytochrome b6f complex decreases with increasing counter-pressure of H\textsuperscript{+}, so at lower pH (Laisk et al., 2005). But
lumenal pH regulates also photoprotection mechanisms through the de-excitation of Photosystem II (PSII) and the activation of non-photochemical quenching (NPQ) under excess light conditions (Trinh and Masuda, 2022). The pH regulation of photoprotection can be finely regulated by H\(^+\)-based exchangers. For example KEA3 has been shown to be localised on thylakoid membrane functioning as K\(^+\)/H\(^+\) antiporter, and influencing NPQ control and pH homeostasis in higher plants (Trinh and Masuda, 2022; Armbruster et al., 2016) but also in diatoms (Seydoux et al., 2022).

Moreover, ΔpH across plastid envelope membranes, caused by the difference in pH between chloroplast stroma (alkaline pH) and cytosol (nearly neutral pH, ~7.1–7.5), may contribute to the transport of ions and metabolites across the chloroplast envelope through H\(^+\)-based exchangers, antiporters, and symporters (Trinh and Masuda, 2022).

If important photosynthetic mechanisms have been found to be regulated by H\(^+\) exchanges between thylakoid membranes, between chloroplast and cytosol there are many other spaces containing proteins needed for cell functioning and intracellular homeostasis. In fact, diatoms own a secondary plastid that derives from a red algal symbiont, and which is limited by four membranes (Whatley and Whatley, 1981). The two outermost membranes are the chloroplast Endoplasmic Reticulum Membrane (cERM), which is continuous with the host outer nuclear envelope, and the Periplastidial Membrane (PPM). The two innermost membranes correspond to the outer and inner Envelope Membranes (oEM and iEM) of the symbiont chloroplast (Flori et al., 2016). Between the PPM and oEM lies a minimised symbiont cytoplasm, the Periplastidial Compartment (PPC). In *Phaeodactylum tricornutum*, PPC-resident proteins are localised in blob-like structures, which remain associated with plastids after cell disruption (Flori et al., 2016; Maier et al., 2015). H\(^+\) and other ions need to be transported from cytosol towards chloroplast and *vice versa* across these four membranes, but still little is known about different proteins, channels and transporters localised there.
Figure 5.1. Diatoms bilobate-shaped secondary plastid surrounded by four membranes. The outermost membrane is in continuum with the host rough endoplasmic reticulum. The space between the two inner- and outermost membrane pairs (PPC) represents the former red algal cytoplasm of the endosymbiont. cERM: chloroplast ER membrane; PPM: periplastidal membrane; PPC: periplastidal compartment; oEM: plastid outer envelope membrane; iEM: plastid inner envelope membrane.

In diatoms most of the plastidial proteins are nucleus encoded and are imported from the cytoplasm. Nucleus encoded proteins delivered to the PPC and the PPM as well as to the stroma are expressed as pre-proteins with a characteristic targeting signal at the N-terminus (Apt et al., 2002; Gruber et al., 2007). It is composed of a signal peptide for co-translational ER import followed by a transit peptide-like sequence, together called Bipartite Targeting Sequence (BTS) (Maier et al., 2022).

The enzymes retrieved in PPM and PPC suggested some interesting hypothesis about the role of these compartments. For example, the Symbiont-specific ERAD (Endoplasmic Reticulum-Associated protein Degradation)-Like Machinery (SELMA) in the PPM and the chaperone Hsp70 in the PPC are needed for protein import and correct folding respectively (Moog et al., 2011). Moreover, the PPC of diatoms is a transition compartment with only low biochemical capacity such as for CO$_2$ concentration mechanisms (CCM), containing two β-type Carbonic Anhydrases, PtCA-I and -II: these enzymes catalyse the reversible interconversion of CO$_2$ and HCO$_3^-$ and are crucial components of the CCM and the CO$_2$ fixation in diatoms (Moog et al., 2011; Maier et al., 2015; Tanaka et al., 2005; Tachibana et al., 2011).

If little is still known about enzymes localised between these membranes, almost nothing is known about transporters localised on them. In particular, ions and molecules need to be transported inside and outside the chloroplast, to allow to photosynthetic apparatus to
respond to intra- and extracellular conditions. Crucially, ions as H\(^+\) and HCO\(_3^-\) need to cross membranes to allow intracellular ion and pH regulation.

In higher plants, several H\(^+\)-dependent transporters, exchangers, and regulators, which are localised on the chloroplast envelope membranes have been shown to play direct roles in stromal pH regulation. Two K\(^+\) efflux antiporters, KEA1 and KEA2, are required for osmotic stress responses and chloroplast development (Tsujii et al., 2019). CHX23 was shown to function as a putative Na\(^+\)-(K\(^+\))/H\(^+\) antiporter for adjusting pH in the cytosol while maintaining alkaline pH in the chloroplast stroma (Song et al., 2004). Other factors that potentially controls H\(^+\) extrusion across the plastid envelope membrane in higher plants are the H\(^+\)-ATPase, acting as an H\(^+\) pump, and the Day-Length-dependent Delayed-Greening 1 transporter (DLDG1), involved in Na\(^+\)/H\(^+\) and/or K\(^+\)/H\(^+\) antiport (Trinh and Masuda, 2022). However, none of these mechanisms regulating pH across membranes, and in particular across chloroplast envelope membranes, have been investigated in diatoms to date.

The family of Low-Affinity Nitrate Transporters, called NPFs, well characterised in many organisms, have been shown to co-transport a broad range of different substrates, beside NO\(_3^-\), together with H\(^+\) or other ions, controlling membranes proton gradient and transporters whose activity is set by intracellular pH sensors (Fan et al., 2017). Diatoms NPFs (diNPFs) diverge into two Clades, one closer to bacteria and the other closer to higher plants (Santin et al., 2021b). However, they are still poorly characterised.

*P. tricornutum*, widely used as model organism for genomic and genetic tools available, owns two NPFs, each belonging to one of the two clades (Santin et al., 2021b). In particular, PtNPF2 belongs to bacteria-like Clade I and it is phylogenetically and structurally closer to bacteria POTs, known to be H\(^+\)-coupled oligopeptide transporters. PtNPF2 showed a differential gene expression in response to light and in different pH conditions (Santin et al., 2021b), encouraging the hypothesis of a possible role in H\(^+\)-coupled transport.

Here, by molecular and functional genetic analyses, new insights were provided on PtNPF2 as a possible H\(^+\) transporter, and on its possible role in controlling intracellular pH homeostasis.

### 5.2. Materials and Methods
5.2.1. Diatom cultures

An axenic culture of *P. tricornutum* Bohlin, CCMP 632 was obtained from the Provasoli-Guillard National Centre for Culture of Marine Phytoplankton. The culture was maintained in autoclaved 0.22 μm F/2 medium without silica (Guillard, 1975), temperature was controlled at 18°C and white fluorescent lights were used at a 12:12 light dark cycle with a photon flux of 90 μmol photons m$^{-2}$ s$^{-1}$. All cultures in this study were grown in the same way under these conditions, except for specific growth curves experiments described below.

5.2.2. PtNPF2 overexpression vectors

*P. tricornutum* cells were transformed with a two-vector system: an overexpression vector containing the PtNPF2 gene fused with a gene encoding for a fluorescent protein and under the control of a strong promoter, and a second vector containing the *Sh-ble* gene under the control of the Light Harvesting Complex *Lhcf6* (before called fucoxanthin-chlorophyll binding protein-F FcpF) promoter, *Lhcf6p*, vector named Lhcf6p-Sh-Ble-Lhcf1t (Falciatore et al., 1999; Zaslavskaia et al., 2001).

The PtNPF2 gene (Phatr3_J47218) was fused to the *YFP* gene at its 3’ end or with the *GFP* gene at its 5’ end. In particular, in the first case, expression of gene fusion was under the control of the Light Harvesting Complex *Lhcf2* (before called fucoxanthin-chlorophyll binding protein-B FcpB) strong promoter, *Lhcf2p*, and the Light Harvesting Complex *Lhcf1* (before called FcpA) terminator, *Lhcf1t*. Cloning was performed using the Gibson Assembly reaction (Gibson et al., 2009, 2010) and NEBuilder Assembly Tool (http://nebuilder.neb.com/) as suggested by manufacturer’s instructions. The pKS-FcpBpAt-C-EYFP vector (Siaut et al., 2007) backbone of 4357 bp containing promoter, 3’ *YFP* gene and terminator was amplified using primers Lhcf2pYPF_for and Lhcf2pYFP_rev, while the full-length PtNPF2 sequence of 1950 bp, without stop-codon, was amplified from genomic DNA using primers Lhcf2p_PtNPF2_for and PtNPF2_YFP_rev (Table 5.1), in order to amplify the full-length sequence with the two tails needed for subsequent assembly. Q5™ High-Fidelity DNA Polymerase was used to amplify the fragments, using the thermal profile [98°C for 30 sec, (98°C for 10 sec, 57°C for 40 sec, 72°C for 2 min) ×35 cycles, 72°C for 2 min]. This construct of 6307 bp, was called Lhcf2p-PtNPF2-YFP-Lhcf1t vector.
For the *GFP* gene at 5’ end of the *PtNPF2* gene, the expression was under the control of the *Pseudo-nitzschia multistriata* histone *H4* constitutive promoter (*PmH4p*). In particular, the *PmH4pH4N-GFP* plasmid (Sabatino et al., 2015) was modified through Gibson Assembly reaction (Gibson et al., 2010, 2009): plasmid backbone of 4457 bp containing promoter, 5’ *GFP* gene without stop-codon and terminator, was amplified using primers *PmH4pGFP_for* and *PmH4pGFP_rev*, and it was assembled with full-length *PtNP2* sequence using *GFP_PtNPF2_for* and *PtNPF2_Lhcf1t_rev* (Table 5.1), so inserting the *PtNPF2* between the *GFP* gene and the *Lhcf1* terminator and creating the 6410 bp *PmH4p-GFP-PtNPF2-Lhcf1t* vector. Also in this case, Q5™ High-Fidelity DNA Polymerase was used to amplify the fragments, using the thermal profile [98°C for 30 sec, (98°C for 10 sec, 57°C for 40 sec, 72°C for 1 min and 30 sec) ×35 cycles, 72°C for 2 min].

**Table 5.1.** List of primers used for plasmid construction.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lhc2pYPF_for</td>
<td>5’-GAATTCGATATCAAGCTTATCGAT-3’</td>
</tr>
<tr>
<td>Lhc2pYFP_rev</td>
<td>5’-GAATTCATGGTGAGCAAGGGCG-3’</td>
</tr>
<tr>
<td>PmH4pGFP_for</td>
<td>5’-GCGGCCGCAACAACTACC-3’</td>
</tr>
<tr>
<td>PmH4pGFP_rev</td>
<td>5’-AGGCCCTTCTTGACAGCTCG-3’</td>
</tr>
<tr>
<td>Lhc2p_PtNPF2_for</td>
<td>5’-atcgataagcttgatagcttgctgATGAGTAGGAGAAGTTCTC-3’</td>
</tr>
<tr>
<td>PtNPF2_YFP_rev</td>
<td>5’-ccctgctgctgctgctgctgctgctgCTC-3’</td>
</tr>
<tr>
<td>GFP_PtNPF2_for</td>
<td>5’-acgagctgctgctgctgctgctgctgCTC-3’</td>
</tr>
<tr>
<td>PtNPF2_Lhcf1t_rev</td>
<td>5’-gaggtatgtgtgtgctgctgctgctgCTC-3’</td>
</tr>
</tbody>
</table>

To perform the Gibson assembly, a total of 0.03 pmols of the backbone and 0.07 pmols of the *PtNPF2* fragment were assembled into a single vector, at 50°C for 1 hour. Chemical transformation of the One Shot™ TOP10 Chemically Competent *E. coli* strain (Invitrogen) was carried out as described in paragraph 4.2.3 “*PtNPFI* overexpression vectors” in Chapter 4. Diagnostic PCR was carried out on clones growing on selective medium, plasmidic DNA was extracted from positive clones, using the GenElute™ Plasmid Maxiprep Kit (Sigma-Aldrich) according to manufacturer’s protocol, and then sequenced.
Biological transformation of wild-type *P. tricornutum* was performed as previously described in Materials and Methods of the Chapter 4 (see paragraph 4.2.4 “Biologist transformation of *P. tricornutum*” in Chapter 4).

### 5.2.3. PCR analysis on transformed cells overexpressing PtNPF2

PCR was performed on genomic DNA from transformants and wild-type cells, picked from selective and control plates. Cells were then resuspended in 20 µL of a Lysis Buffer consisting of 1% (v/v) Triton X, 20 mM Tris-HCl (pH 8) and 2 mM EDTA (pH 8) in Milli-Q water, vortexed, put on ice for 15 minutes and then at 85°C for 10 minutes, as described in Daboussi et al. (2014). The diluted solution was then used for PCR screening.

To analyse *NPF2-YFP* cassette integration in genomic DNA, the thermal profile used was [95°C for 1 min, (95°C for 20 sec, 58°C for 20 sec, 72°C for 2 min and 30 sec) × 35 cycles, 72°C for 2 min], using primers Lhcf2p_for and PtNPF2_exp_rev (fragment 760 bp), and PtNPF2_exp_for and GFP_rev (fragment 2247 bp) (Table 5.2).

To analyse *GFP-NPF2* cassette integration in genomic DNA, the thermal profile used was [95°C for 1 min, (95°C for 20 sec, 62°C for 20 sec, 72°C for 2 min) × 35 cycles, 72°C for 2 min] with primers PmH4p1_for and GFP_rev (fragment 1319 bp), and [95°C for 1 min, (95°C for 20 sec, 58°C for 20 sec, 72°C for 2 min and 30 sec) × 35 cycles, 72°C for 2 min] with primers GFP_for and PtNPF2_exp_rev (fragment 2128 bp) (Table 5.2). Amplification fragments were detected on agarose gel.

### Table 5.2. List of primers used for overexpressing mutants screening.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lhcf2p_for</td>
<td>5'-CGCCGTAAACAGCAAATCCT-3’</td>
</tr>
<tr>
<td>PmH4p1_for</td>
<td>5'-AACAAAGCTGGGTACCGGC-3’</td>
</tr>
<tr>
<td>PtNPF2_exp_for</td>
<td>5’-TTACGTGATGGGCTTGCTCCA-3’</td>
</tr>
<tr>
<td>PtNPF2_exp_rev</td>
<td>5’-GGTCCGGCGGTTATTACACGA-3’</td>
</tr>
<tr>
<td>GFP_for</td>
<td>5’-CCAGCAGAACACCCCAT-3’</td>
</tr>
<tr>
<td>GFP_rev</td>
<td>5’-AACTCCAGCAGGACCATGTG-3’</td>
</tr>
</tbody>
</table>
5.2.4. RNA extraction and qPCR analysis

RNA extraction and quantitative Real-Time reverse-transcription Polymerase Chain Reaction (qRT-PCR or qPCR) were performed as described in Russo et al. (2015). Around 50 ml of cells growing in exponential phase (1.5 - 2 \cdot 10^6 cell/ml) were collected by centrifuging for 10 minutes at 4000 g at 4 °C, and then extracted with Trizol™ (Invitrogen), as previously described in paragraph 2.2.6 “RNA extraction, primer design and qPCR” in Chapter 2.

RNA quantity was determined using Qubit™ 2.0 Fluorometer (Life Technologies), while RNA quality was checked through 1% agarose gel electrophoresis. 200 ng of total RNA were retro-transcribed using the PrimeScript™ RT Reagent Kit (Perfect Real Time) (TakaraBio). To assess cDNA quality and absence of gDNA contamination, Reverse Transcriptase (RT) -PCRs were run with control PtNOA gene (Nitric Oxide Associated protein, Phat3_J40200): intron spanning primers PtNOA_for and PtNOA_rev were used with the thermal profile [95°C for 1 min, (95°C for 20 sec, 57°C for 20 sec, 72°C for 30 sec) \times 35 cycles, 72°C for 2 min].

1 μl of a 1:2 dilution of cDNA was used as template to amplify the transcripts using 0.4 μM final concentration of the primers. The RPS gene (Ribosomal Protein Small Subunit 30S, Phatr3_J10847) was used as reference gene for qPCR data normalization (Siaut et al., 2007). Genes and primers used are reported in Table 5.3.

Table 5.3. List of primers used for qPCRs.

<table>
<thead>
<tr>
<th>Target gene</th>
<th>Gene ID</th>
<th>Primer name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PtNOA</td>
<td>40200</td>
<td>PtNOA_exp_for</td>
<td>5’-CAGTTACTGACCCCCGAAGA-3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PtNOA_exp_rev</td>
<td>5’-AACGCACTTTCCGAGAAGAG-3’</td>
</tr>
<tr>
<td>PtRPS</td>
<td>10847</td>
<td>PtRPS_for</td>
<td>5’-GTGCAAGAGACCAGATGACC-3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PtRPS_rev</td>
<td>5’-CGAAGTCAACCCAGAAACC-3’</td>
</tr>
<tr>
<td>PtNPF2</td>
<td>47218</td>
<td>PtNPF2_exp_for</td>
<td>5’-TTACGATGATGCTTTGCTCCA-3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PtNPF2_exp_rev</td>
<td>5’-GGTCCGGCGTATAACAGA-3’</td>
</tr>
</tbody>
</table>

qPCR amplification was performed using Fast SYBR Green Master mix with ROX (Applied Biosystems) in a final volume of 10 μL, using ViiA™ 7 Real-Time PCR System (Applied Biosystems). qPCR conditions used were as follows: 95 °C for 20 sec, 40 cycles
at 95 °C for 1 sec and 60 °C for 20 sec, 95 °C for 15 sec, 60 °C 1 min, and a gradient from 60 °C to 95 °C for 15 min.

Data obtained were processed with the ViiA™ 7 Real-Time PCR system software. Fold-changes were obtained with the Relative Expression Software Tool-Multiple Condition Solver (REST-MCS) (Pfaffl et al., 2002) and values above ±2 were considered significant.

5.2.5. Confocal Microscopy

Images of wild-type *P. tricornutum* cells and *PtNPF2* overexpressing ones were recorded on a Leica SP8 X confocal laser-scanning microscope. All images were collected using the HC PL APO CS2 63X/1.20 water objective.

Excitation of GFP and YFP fluorescence was performed at 488 nm using a 65-mW argon laser and detected at 500-562 nm, while chlorophyll *a* autofluorescence was observed through excitation at 554 nm and detection of fluorescence emission at 650-741 nm. Hoechst 33342 (Life Technologies) was used at a final concentration of 5 μg/ml to stain nuclear DNA; stained cells were then visualised by excitation at 405 nm and detection at 424-462 nm.

5.2.6. Growth curves in different conditions

*P. tricornutum* wild-type, *PtNPF2-YFP* overexpressing strains 2OE 2 and 2OE 3, and *Ptnpf2* knock-out lines 2KO 1.15 and 2KO 1.16 were grown axenically in F/2 medium without silica (Guillard, 1975) to mid-exponential phase and then transferred to different external medium conditions to perform growth curve experiments. A knock-out mutant for a different gene, generated with the same proteolistic protocol (1KO 2.9), was used as a second control strain. All growth curves were performed at 18°C and white light was provided at a 12:12 light dark cycle with a photon flux of 90 μmol photons m⁻² s⁻¹.

In detail, cells were grown in triplicate in vented cap flasks in different NO₃⁻ concentrations for 10 days: F/2 medium without silica with 882 μM NaNO₃, set as reference condition, or 50 μM NaNO₃, set as N starvation condition. Then, cells were grown in triplicate in vented cap flasks with different N sources for 10 days: F/2 medium without silica with 882 μM NaNO₃, 882 μM NH₄Cl or 882 μM urea. Another experiment consisted in cells growing in triplicate in vented cap flasks in different medium pH for 10 days: pH medium was calibrated at the beginning of the growth curve, at pH 6, pH 7, pH
8 set as normal condition, and pH 9, through using HCl 6N. During these growth curves, cell concentration was evaluated using a Malassez cell counting chamber, and in vivo chlorophyll a fluorescence was measured as a proxy for growth (Supplementary Fig. S4.1), using multifunctional monochromator-based microplate reader (Infinite™ M1000 Pro; Tecan), with excitation and emission wavelengths set at 662 and 685 nm, respectively (Ni et al., 2019; Mandalakis et al., 2017; Kalaji et al., 2014).

5.2.7. Normal to low pH shift

Normal to low pH shift experiment was performed on P. tricornutum wild-type, PtNPF2-YFP overexpressing strains 2OE 2 and 2OE 3, and Ptnpf2 knock-out lines 2KO 1.15 and 2KO 1.16. Cells were grown at 18°C, 12:12 light dark cycle under white light with a photon flux of 90 μmol photons m⁻² s⁻¹. Cells growing in Erlenmeyer flasks at the exponential phase (1.5-2 · 10⁶ cells/ml) in F/2 medium without silicates at pH 8 (normal pH), were diluted 1:10 in F/2 medium without silicates at pH 7 (low pH). The experiment was performed in triplicate.

Cells were counted and photosynthetic parameters and relative intracellular pH were measured just before the shift, and 3, 24 and 48 hours after the shift to pH 7 (Fig. 5.2). Cell concentration was evaluated through in vivo chlorophyll a fluorescence using multifunctional monochromator-based microplate reader (Infinite™ M1000 Pro; Tecan), and confirmed by using a Z2 Coulter Particle Count and Size Analyser (Beckman Coulter). Minimal and maximal size thresholds were set respectively to 3 and 8 μm. Then samples were collected for subsequent measurements (Fig. 5.2).

![Figure 5.2](image.png)

**Figure 5.2.** Pipeline of the shift experiment from normal pH (pH 8) to low pH (pH 7), including the list of P. tricornutum strains used, the time points analysed and the measurements performed. The three Erlenmeyer flasks represent three biological replicates for each strain.
5.2.8. Dark to light shift

Dark to light shift experiment was performed on *P. tricornutum* wild-type, *PtNPF2-YFP* overexpressing strains 2OE 2 and 2OE 3, and *Ptnpf2* knock-out lines 2KO 1.15 and 2KO 1.16. Cells growing in Erlenmeyer flasks at the exponential phase in F/2 medium without silicates, with a photon flux of 90 μmol photons m\(^{-2}\) s\(^{-1}\), were diluted to 1 · 10\(^{6}\) cells/ml and put in dark for four days. Dark condition was applied though covering Erlenmeyer flasks with aluminium foil avoiding light exposure. After four days, aluminium foils were removed to re-illuminate cells. The experiment was performed in triplicate. Just before the shift (dark), 3 (3h light) and 24 hours after shift to illumination (24h light), cell concentration was evaluated through *in vivo* chlorophyll \(a\) fluorescence using multifunctional monochromator-based microplate reader (Infinite™ M1000 Pro; Tecan), and confirmed by using a Z2 Coulter Particle Count and Size Analyser (Beckman Coulter). Minimal and maximal size thresholds were set respectively to 3 and 8 μm. Then samples were collected for subsequent photosynthetic parameters measurements.

5.2.9. Photosynthetic measurements

All fluorescence analyses were performed on cultures grown in liquid F/2 medium without silica. To measure Photosystem II (PSII) parameters, 10 to 50 ml of biological replicate cultures (to have around 10\(^{6}\) total cells) were centrifuged at 2500 g for 15 minutes and then concentrated ten times in their own supernatant. Concentrated cells were adapted to low light for 20 minutes to induce complete relaxation of the electron transport chain. Fluorescence analysis was conducted using a fluorescence CCD camera recorder (SpeedZen, JBeamBio, France). Measuring light and actinic saturating flash were provided by blue and green LEDs, respectively. Maximum quantum yield of PSII in the dark was calculated as \(\frac{F_v}{F_m} = \frac{(F_m-F_0)}{F_m}\), where \(F_0\) is the minimum fluorescence of the dark-adapted (20 minute) sample probed by a non-actinic detecting light, and \(F_m\) the maximum fluorescence detected immediately after a saturating light pulse (250 ms, 532 nm, 5000 μmol photons m\(^{-2}\) s\(^{-1}\)) is applied on the dark-adapted sample (Genty et al., 1989).

Samples were then exposed to increasing intensity steps of green actinic light provided by LEDs, to construct the saturation light curves (Steady State Light Curves – SSLCs). Each light step lasted at least five minutes to ensure full adaptation of the photosynthetic apparatus to the irradiation level. At the end of each light step, steady-state minimum and maximum fluorescence of PSII under light (\(F_s\) and \(F_m'\), respectively) were measured.
PSII quantum yield $\Phi_{\text{PSII}}$ was calculated as $(F_{m}^\prime-F_s)/F_m^\prime$ (Maxwell and Johnson, 2000). For each light step, the relative Electron Transfer Rate through PSII (rETR) was calculated as $\Phi_{\text{PSII}} \times$ Light intensity ($E$). Relative ETR (rETR)-photosynthetic light intensity curves were fitted by $\text{rETR} = \text{rETR}_{\text{max}} \cdot (1-\exp(-\alpha \cdot E)/\text{rETR}_{\text{max}})$ as in Blommaert et al. (2021), whereby $\alpha$ is the initial slope of light-limited rETR and rETR$_{\text{max}}$ the saturated rETR value. $E_k$, the light saturation parameter, is calculated as rETR$_{\text{m}}$/\alpha (Webb et al., 1974).

Non-Photochemical Quenching (NPQ) was calculated as $(F_m^\prime-F_s^\prime)/F_m^\prime$, where $F_m$ is the maximal fluorescence of dark-adapted samples with all PSII centres closed and $F_m^\prime$ is the maximal fluorescence of illuminated samples (Maxwell and Johnson, 2000). NPQ values were obtained at the last 750 µmol photons m$^{-2}$ s$^{-1}$ light step. NPQ-photosynthetic light intensity curves were fitted by $\text{NPQ} = \text{NPQ}_{\text{max}} \cdot E^n/(E_{50}^{\text{NPQ}} + E^n)$, whereby $E_{50}^{\text{NPQ}}$ is the light intensity for which 50% of the predicted maximum NPQ value (NPQ$_{\text{max}}$) is reached and $n$ represents the sigmoidicity parameter of the function (Serôdio and Lavaud, 2011; Blommaert et al., 2021).

After the SSLC curve, during the final dark period, $F_0^\prime$ was continuously monitored and the kinetics of fluorescence recovery and NPQ relaxation determined through frequent determinations of $F_m$. All the parameters measured have been summarised in the Supplementary Figure S5.1.

5.2.10. Mini-FIRe measurements

Variable fluorescence was also measured with a miniaturised Fluorescence Induction and Relaxation Systems (called mini-FIRe) developed by (Gorbunov et al., 2020). The peak excitation intensity in the measuring volume of the FIRe instrument was optimally adjusted to ensure closure of PSII reaction centres and saturation of fluorescence within approximately 100 µs (i.e., a single electron turnover in PSII) (Gorbunov et al., 2020).

The functional absorption cross-section of PSII antenna ($\sigma$) is the product of the optical absorption cross-section of PSII (i.e., the physical size of the PSII antennae) and the quantum yield of photochemistry in PSII (Falkowski et al., 2004b). $\sigma$ was calculated from the fluorescence induction upon a PSII single turnover flash of blue light (100 µs, 455 nm, 60 nm bandwidth). The induction curves were measured on 20 minutes dark-acclimated samples without centrifugation with the mini-FIRe fluorometer (Gorbunov and Falkowski, 2021).
The mini-FIRE measures minimum and maximum fluorescence, $F_0$ and $F_m$. $F_m$ was measured as proxy for growth. Moreover, it has been proposed that the relative number of functional PSII reaction centres (RCII) is approximately proportional to fluorescence emitted from open RCII divided by the absorption cross section of PSII photochemistry (RCII=$F_0/\sigma$) (Oxborough et al., 2012). All the parameters measured have been summarised in the Supplementary Figure S5.1.

5.2.11. Intracellular pH measurements

The relative intracellular cytosolic pH value was measured mainly following Dixon et al. (1989) and Shi et al. (2019). $4 \cdot 10^6$ total cells of *P. tricornutum* were centrifuged for 10 minutes at 3600 g at 18°C and then carefully resuspended into 400 µl PBS (1X). A 5 µM final concentration of BCECF/AM (Molecular Probe Life Technologies, Carlsbad, CA, USA), previously re-suspended in DMSO for a stock solution of 1 mM, was then added. After 30 minutes of incubation at 37°C, cells were centrifuged and washed three times with PBS (1X). Fluorescence emission at the wavelength of 530 nm was measured at the excitation wavelengths of 420 nm and 470 nm (the isosbestic point of the dye) though using the Infinite® M1000Pro Plate Reader (TECAN).

The ratio of the emission with excitation at 470 nm to that at 420 nm was calculated and is proportional to the intracellular pH. The relative intracellular pH was then calculated by normalising the ratio obtained with treated cells to that with control cells, generally wild-type at time zero (Shi et al., 2019).

5.2.12. Bicarbonate assay

Exponentially growing cultures of *P. tricornutum* wild-type, *PtNPF2-YFP* overexpressing strains 2OE 2 and 2OE 3, and *Ptnpf2* knock-out lines 2KO 1.15 and 2KO 1.16, were washed three times in HCO$_3^-$-free artificial sea water medium buffered with 10 mM HEPES (pH 8) and resuspended in 500 µl at a concentration of ~2 · 10$^7$ cell/ml. Oxygen evolution measurements were performed using a Clack’s-type dissolved oxygen (DO) cuvette electrode (Qubit Systems Inc., ON, Canada). Cells were illuminated at 240 µmol photons m$^{-2}$ s$^{-1}$ and oxygen production was measured after the addition of increasing amounts of HCO$_3^-$. Maximum rate of net photosynthesis corresponds to the point where oxygen evolution remains unchanged regardless of the Dissolved Inorganic Carbon (DIC) concentration.
5.2.13. Statistical analyses

To determine whether there were significant differences in growth, photosynthetic parameters and intracellular pH, between the different strains and over time, a two-way ANOVA with Tukey’s Multiple Comparison Test was performed, and statistics for time, strain, and interaction were recorded. Pairwise post-hoc Tukey tests were applied for all multiple comparison procedures when differences were found significant ($p < 0.05$). All statistical analyses were performed using GraphPad Prism version 6.0.

5.2.14. Sample preparation for RNA-seq

Cells shifted from normal to low pH were collected at two time points: just before the shift (exponential phase at pH 8) and 24 hours after the shift to low pH (exponential phase at pH 7). Around $10^8$ cells were collected by centrifuging for 10 minutes at 3500 g at 4 °C, and extracted with Trizol™ (Invitrogen), as previously described in paragraph 2.2.6 “RNA extraction, primer design and qPCR” in Chapter 2.

RNA was then purified by an ammonium acetate (NH$_4$Ac) precipitation to remove impurities: one volume of NH$_4$Ac 5M was added to RNA, then it was mixed, incubated for 15 minutes on ice, centrifuged for 15 minutes at 14 000 g at 4°C. The pellet was washed with 1 ml ethanol 70%, dried and resuspended in RNase-free water.

Subsequently, RNA quantity was determined using Qubit™ 2.0 Fluorometer (Life Technologies), while RNA quality was checked through 1% agarose gel electrophoresis. DNase I (Qiagen) treatment was applied to remove the gDNA contamination, and RNA was further purified using Direct-zol™ RNA Microprep (ZYMO Research). RNA quantification was done again using Qubit™ 2.0 Fluorometer (Life Technologies) and integrity was assessed using Bioanalyzer (2100 Bioanalyzer Instruments, Agilent Technologies).

5.2.15. Vector for PtNPF2 heterologous expression in Xenopus laevis

The vector pGEM-Xho, a modified pGEM vector kindly provided by Dr. Lacombe (Institut des Sciences des Plantes de Montpellier - France), and containing the gene for ampicillin resistance, was used as backbone for cloning PtNPF2 full-length sequence. PtNPF2 was amplified from genomic DNA, using primers PtNPF2_EcoRI_for 5’-ccgGAATTCATGCCGAGGCACC-3’ and PtNPF2_XbaI_rev 5’-ctagTCTAGATCAATCATTCTTCGCTCTTC-3’, with underlined restriction site.
The PtNPF2 full-length sequence and vector pGEM-Xho were both cut with restriction enzymes EcoRI and XbaI, assembled by using T4 ligase and transformed in the One Shot™ TOP10 Chemically Competent *E. coli* strain (Invitrogen), as described in the paragraph 4.2.12 “Vector for PtNPF1 heterologous expression in *Xenopus laevis*” in Chapter 4.

Diagnostic digestion was carried out on clones growing on selective medium and plasmid DNA was then extracted from positive clones, using the GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich) and according to manufacturer’s protocol. Plasmidic DNA was sequenced, quantified using Qubit™ 2.0 Fluorometer (Life Technologies), and sent to Dr. Lacombe at the Institut des Sciences des Plantes de Montpellier (France) for the heterologous expression in *Xenopus laevis* oocytes.

### 5.3. Results

#### 5.3.1. Generation and selection of *P. tricornutum* PtNPF2 overexpressing strains

Similarly to PtNPF1 (see paragraph 4.3.2 “Generation and selection of *P. tricornutum* PtNPF1 overexpressing strains” in Chapter 4), co-transformation was performed in order to tag PtNPF2 with a fluorescent protein and determine its subcellular localisation.

*P. tricornutum* cells were transformed with a plasmid containing the *Sh-ble* resistance gene and a second plasmid containing the full-length PtNPF2 sequence. In detail, for PtNPF2 two different plasmids were generated: the first one with the PtNPF2 gene under the control of the strong *Lhcf2p* promoter and in frame with a 3’ YFP, while the other under the control of the constitutive *PmH4p* promoter and fused with a 5’ GFP. The generation of both constructs allowed to make the fluorescence-based localisation more reliable. Plasmids used for biolistic transformations are shown in Figure 5.3.
Figure 5.3. Vectors generated for PtNPF2 overexpression, through co-transformation with resistance plasmid: A) Lhcf2p-PtNPF2-YFP-Lhcf1t plasmid, with PtNPF2 upstream of YFP and controlled by the *P. tricornutum* Lhcf2 strong promoter; B) PmH4p-GFP-PtNPF2-Lhcf1t plasmid, with PtNPF2 downstream of GFP and controlled by the *P. multistriata* PmH4 constitutive promoter; C) Lhcf6p-Sh-Ble-Lhcf1t plasmid, with *Sh-Ble* gene conferring Phleomycin resistance (Falciatore et al., 1999).

After three weeks from biolistic transformation, colonies positive to phleomycin selection and non-transformed wild-type control cells were analysed through PCR, using two pairs of primers to check the whole cassette insertion in *P. tricornutum* genome (Fig. 5.4A-B).

As regards to *PtNPF2-YFP* overexpressing strains, among 12 positive colonies analysed through PCR, 7 show the whole *PtNPF2-YFP* cassette insertion, so around 60% of the screened clones (Fig. 5.4C).
**Figure 5.4.** PCR screening on *P. tricornutum* colonies growing on selective medium. A) Transformed cells growing on selective medium, with non-transformed control on selective (C-) and non-selective (C+) plates. B) Schematic representation of the PtNPF2-YFP cassette with the two primer couples used, represented by red and green arrows and amplifying two partially overlapping regions. C) PCR showing amplification of a fragment from promoter to the gene of interest PtNPF2 and a fragment from PtNPF2 to YFP. M represents the 1Kb marker, C+ the plasmid and C- the blank.

For *PtNPF2-YFP* overexpressing clones 2OE 2 and 2OE 3, which showed the whole cassette inserted (Fig. 5.4C), the *PtNPF2* gene expression levels were analysed. 2OE 2 shows a 6.87 ±0.42 -fold increase, while 2OE 3 shows a 6.39 ±0.36 -fold increase, compared to wild-type (Fig. 5.5).
Figure 5.5. *PtNPF2* relative gene expression levels of *P. tricornutum PtNPF2* overexpressing strains 2OE 2 and 2OE 3, normalised on the internal control *RPS* gene and on the wild-type, set as zero. Error bars represent the standard deviation of three technical replicates of two biological replicates (for a total of 6 measurements).

Also, *GFP-PtNPF2* overexpressing strains were screened using two primer couples to check the correct cassette integration (Fig. 5.6A): among 10 positive colonies analysed through PCR, 4 show the whole *GFP-PtNPF2* cassette insertion, so around 40% of the screened clones (Fig. 5.6B).
Figure 5.6. PCR screening on *P. tricornutum* colonies growing on selective medium. A) Schematic representation of the GFP-PtNPF1 cassette with the two primer couples used, represented by red and green arrows and amplifying two partially overlapping regions. B) PCR showing amplification of a fragment from promoter to the GFP gene and a fragment from GFP to the gene of interest PtNPF2. M represents the 1Kb marker, C+ the plasmid and C- the blank.

5.3.2. *In vivo* subcellular localisation of PtNPF2

Subcellular localisation of PtNPF2 was determined by observing PtNPF2 overexpressing strains through confocal microscopy.

First, *in vivo* localisation studies were performed on the two detected candidates expressing PtNPF2 with C-terminal YFP, 2OE 2 and 2OE 3. Both clones selected show the same localisation pattern, with the fusion protein fluorescence localised in a small blob-like structure close to the nucleus (Fig. 5.7). In some cases, this structure is a single spot while in other ones it has a bilobate shape. Its position is between the nucleus, with which it does not co-localise, and the chloroplast, generally in the central part of the cell. In particular, the one or two dot-like fluorescence structures are near the middle of the plastid, where the lobes of the autofluorescence are constricted. In fact, the blob-like structure is due to a median constriction of the two innermost membranes of the plastid (the plastid envelope membranes oEM and iEM), which leads to a widening of the PPC (Flori et al., 2016). For its specific shape and position, PtNPF2 fused with fluorescent
tags is hypothesised to be localised on the PPC which is surrounded by the PPM (Moog et al., 2011; Marter et al., 2020).

**Figure 5.7.** Subcellular localisation of PtNPF2 in *P. tricornutum* through PtNPF2-YFP fusion protein expression. First column: BF corresponds to Bright-Field; second column: chlorophyll *a* autofluorescence in red; third column: YFP fluorescence in green; fourth column: HOECST staining for nucleus visualisation in blue; fifth column: merged channels. Scale bar: 5 µm.

Then, fluorescent GFP-PtNPF2 signal was observed through confocal microscopy, showing the same pattern previously observed in *PtNPF2-YFP* overexpressing strains and confirming the PtNPF2 subcellular localisation. The small green dot has a central position inside the cell, close to the chloroplast and sometimes forming two lobes (Fig. 5.8).
In order to confirm that the subcellular localisation is on the PPC compartment, PtNPF2-YFP overexpressing strain 2OE 3 was further transformed with pPha-NR-sHsp70_BTS_mRuby3-NAT plasmid (Fig. 5.9), kindly provided by Dr. Meier at the University of Marburg (Germany) (Marter et al., 2020). In this plasmid, the gene encoding mRuby3 was adapted to the codon usage of *P. tricornutum* and located downstream of the bipartite targeting sequence (BTS) of the gene encoding a Heat-Shock Protein of 70 kiloDalton (Hsp70), a chaperone experimentally localised to the PPC of *P.*
tricornutum. This cassette was inserted into a modified version of the pPha-NR shuttle vector (NCBI accession number: JN180663), in which the resistance gene for Phleomycin (Sh-ble) was replaced by a gene conferring resistance against Nourseothricin (nat, pPha-NR-NAT) (Marter et al., 2020). P. tricornutum wild-type and PtNPF2-YFP overexpressing strain 2OE 3 were transformed with this plasmid, called pPha-NR-sHsp70_BTS_mRuby3-NAT, through biolistic protocol. Cells were replated on 300 µg/ml Nourseothricin and positive colonies are expected to grow in the next weeks.

Figure 5.9. pPha-NR-sHsp70_BTS_mRuby3-NAT plasmid provided by Dr. Meier, containing the BTS of the gene encoding Hsp70 in frame with the mRuby fluorescent tag and the gene conferring resistance against Nourseothricin.

5.3.3. Phenotypic analyses of mutants

Wild-type, PtNPF2-YFP overexpressing strains 2OE 2 and 2OE 3, and Ptnpf2 knock-out mutants 2KO 1.15 and 2KO 1.16 (knock-out mutants generation has been described in paragraph 3.3.6 “Ptapt-Ptnpf2 knock-out mutants generation and screening” in Chapter 3 and summarised in Figure 5.10) were used to perform different growth curves and experiments.
Figure 5.10. *PtNPF2* gene schemes of *P. tricornutum* wild-type and *Ptnpf2* mutants 2KO 1.15 and 1.16, showing Cas9 different cuts on the two mutated alleles. Yellow bars represent the two gRNAs.

First of all, PtNPFs were classified as putative NO$_3^-$ transporters, but from previous gene expression analyses no differential *PtNPF1* gene expression was observed in wild-type cells growing in different NO$_3^-$ concentrations or N sources (see paragraph 2.3.5 “Expression patterns of NPFs in *P. tricornutum*” in Chapter 2). This is possible as, in other organisms such as higher plants, NPFs do not transport only NO$_3^-$ but also a big range of other molecules (Léran et al., 2014). To confirm that environmental N availability does not influence PtNPF1 function, wild-type and transgenic *P. tricornutum* lines were grown in two different NO$_3^-$ concentrations, 882 µM set as normal condition and 50 µM set as N starvation condition: no difference in growth has been observed in different strains (Figs. 5.11C-D). Moreover, different N sources were tested, in particular the same concentration of NO$_3^-$, set as reference condition, NH$_4^+$ and urea. Generally, cells grow faster in NH$_4^+$ than NO$_3^-$ and urea, consistently with the fact that NH$_4^+$ is a more cost-effective N source to exploit from an energy point of view. However, no differences have been observed between different strains (Figs. 5.11E-H).
Figure 5.11. Growth curves of *P. tricornutum* wild-type, two *PtNPF2-YFP* strains and two *Ptnpf2* mutants in different NO$_3^-$ concentrations and N sources: A-B) 882 µM NaNO$_3$, C-D) 50 µM
NaNO₃, E-F) 882 µM NH₄Cl and G-H) 882 µM urea. In vivo chlorophyll a fluorescence was used for measurements as proxy for growth, and natural logarithm of growth curves was calculated. A knock-out mutant for a different gene, generated with the same proteolistic protocol (1KO 2.9), was used as additional control strain. In all the graphs, error bars represent the standard deviation of three biological replicates. a.u. arbitrary units. * indicates $p < 0.05$; ** for $p < 0.01$ and *** for $p < 0.001$.

As PtNP2 showed a differential regulation at different pH conditions (see paragraph 2.3.5 “Expression patterns of NPFs in P. tricornutum” in Chapter 2), suggesting its involvement in H⁺ transport, P. tricornutum strains were grown in different pH conditions: low (pH 7), normal (pH 8) which corresponds to normal condition and high pH (pH 9). A reduction in Ptnpf2 mutants cell concentration has been observed at low pH (Figs. 5.12E-F), compared to wild-type and PtNP2-YFP overexpressing strains, reduction not observed in the higher pH conditions (Figs. 5.12A-D). In particular, observing the growth curve at logarithmic scale (Figs. 5.12F), Ptnpf2 mutants and wild-type growths are parallel from day 2, suggesting that the phenotype is a lag in growth after the change of pH. The same growth curve was also performed at very-low pH (pH 6): similarly to pH 7, a lag in growth of both Ptnpf2 mutants has been observed (Figs. 5.12G-H). However, pH decrease affects the growth of all the strains, so that the difference observed at pH 6 is smaller than that seen at pH 7, probably because of the pleiotropic effect caused by the excessive pH decrease.
Figure 5.12. Growth curves of *P. tricornutum* wild-type, two *PtNPF2-YFP* overexpressing strains and two *Pmpf2* mutants in different pH conditions: A-B) pH 9, C-D) pH 8, E-F) pH 7 and G-H) pH 6. *In vivo* chlorophyll a fluorescence was used for measurements as proxy for growth.
and natural logarithm of growth curves was calculated. A knock-out mutant for a different gene, generated with the same proteolistic protocol (1KO 2.9), was used as additional control strain. In all the graphs, error bars represent the standard deviation of three biological replicates. a.u. arbitrary units. * indicates $p < 0.05$; ** for $p < 0.01$ and *** for $p < 0.001$.

Combination of low pH and different N sources was also used to perform growth curves on different *P. tricornutum* strains. NO$_3^-$, set as reference condition, NH$_4^+$ and urea were used as different N sources for medium at pH 7. This experiment was performed to test a possible correlation between H$^+$ and NO$_3^-$ transport. A difference in cell concentration between wild-type and *Ptnpf2* knock-out mutants has been observed in all the conditions (Fig. 5.13), supporting once again the hypothesis that environmental N source could not be responsible for the growth phenotype observed, and a link between H$^+$ and NO$_3^-$ transport could be excluded.
Figure 5.13. Growth curves of *P. tricornutum* wild-type, two *PtNPF2-YFP* overexpressing strains and two *PtNpf2* mutants in medium at pH 7 supplemented with different N sources: A-B) 882 µM NaNO$_3$, C-D) 882 µM NH$_4$Cl and E-F) 882 µM urea. *In vivo* chlorophyll *a* fluorescence was used for measurements as proxy for growth, and natural logarithm of growth curves was calculated. A knock-out mutant for a different gene, generated with the same proteolistic protocol (1KO 2.9), was used as additional control strain. In all the graphs, error bars represent the standard deviation of three biological replicates. a.u. arbitrary units. * indicates $p < 0.05$; ** for $p < 0.01$ and *** for $p < 0.001$. 
5.3.4. Normal to low pH shift

More investigations were done on the transient growth reduction observed in the two Ptnpf2 mutants at pH 7 (Fig. 5.12E-F) compared with other strains. This reduction, evident also from the paler colour of Ptnpf2 mutants cultures (Fig. 5.14A), is mainly due to a low growth rate in the first 24 hours, of 0.44 ±0.18 and 0.48 ±0.07 in Ptnpf2 knock-out strains 1.15 and 1.16, respectively, compared to 0.80 ±0.08 in wild-type (Fig. 5.14B). However, Ptnpf2 mutants recover in the next days, reaching growth rates similar to those of the other strains (Fig. 5.14), and confirming the transient phenotype.

Figure 5.14. Growth of P. tricornutum wild-type, PtNPF2-YFP overexpressing strains and Ptnpf2 mutants. A) Photo of different wild-type and transgenic strains after four days from the shift to low pH (pH 7). Labels indicate different strains. Note the paler colour of flasks of the Ptnpf2 knock-out mutants. B) Growth rates calculated on cell concentrations of P. tricornutum wild-type, two PtNPF2-YFP overexpressing strains and two Ptnpf2 mutants, calculated at days 1 and 2 on previous growth curves performed at pH 7 (Figs. 5.12E-F). Error bars represent standard deviations of three biological replicates. * indicates $p < 0.05$. 
In the context of an EMBO Scientific Exchange Grant, a short-term experiment was performed, trying to amplify the transient differences observed in the first part of the growth curve.

Wild-type, two *PtNP2-YFP* overexpressing strains and two *Ptnpf2* knock-out mutants were shifted from normal (pH 8) to low pH (pH 7), and also to normal pH as control. Before the shift and after 3, 24 and 48 hours, cells were counted, and photosynthetic parameters and relative intracellular pH were measured.

After 3 hours from the shift a strong reduction in growth of all the strains shifted to low pH has been observed, compared to the same strains growing at normal pH, probably because of the pH stress introduced in cell cultures (Fig. 5.15). Then, a significant growth reduction of around 25% in the first 24 hours of the two *Ptnpf2* knock-out strains has been confirmed (Figs. 5.15B, D, F and S5.1B), with a growth rate recovery observed the next day, when all the strains reach similar growth rates (Fig. 5.15F).
Figure 5.15. Growth curves of *P. tricornutum* wild-type, two *PtNPF2-YFP* overexpressing strains and two *Pttnpf2* mutants after the shift to A) pH 8 and B) pH 7. Natural logarithm of growth curves was calculated, respectively for shift to C) pH 8 and D) pH 7. Growth rates were calculated at different time points for the shift to E) pH 8 and F) pH 7. Error bars represent standard deviations of three biological replicates. * indicates $p < 0.05$; ** for $p < 0.01$ and *** for $p < 0.001$.

This transient phenotype was then further investigated in terms of photosynthetic performances and intracellular pH.

The maximum dark-adapted yield of PSII ($F_v/F_m$), is lower in *Pttnpf2* knock-out mutants compared with other strains at all the time points after the shift to low pH (Fig. 5.16B). By contrast, no differences in $F_v/F_m$ have been observed between strains after the shift to normal pH (Fig. 5.16A). The lower $F_v/F_m$ in *Pttnpf2* mutants corroborates physiology...
disruption consistent with the growth reduction observed at low pH, however PSII efficiency needs to be probed under light to make a direct link with growth (see below).

Then, PSII driven relative Electron transport Rate (rETR) and photoprotection via Non-Photochemical Quenching (NPQ) were studied: all the analysed parameters are summarised in Supplementary Figures S5.1A and B.

Lower maximum rETR (rETR\textsubscript{max}) has been observed in \textit{PtNPF2} knock-out mutants compared to wild-type after 3 and 24 hours from the shifts (Fig. 5.17), with the strongest difference observed just after 3 hours from the shift to pH 7. In particular, this difference is greater after the shift to low pH than to normal pH, with a rETR\textsubscript{max} significantly lower, about 13-18\%, in \textit{PtNPF2} strains after 3 hours and of 5-12\% after 24 hours, compared to wild-type (Figs. 5.17D-E). Moreover, after 48 hours following the shift to normal pH, a
higher rETR\textsubscript{max} has been observed in overexpressing strains (Fig. 5.17C). On the other hand, no differences have been observed in \( \alpha \), the initial slope of the ETR curve which represents the efficiency of electron transport at non-saturating lights, and in the light saturation index \( E_k \) (Supplementary Fig. S5.3).

**Figure 5.17.** rETR\textsubscript{max} values for *P. tricornutum* wild-type, two *PtNPF2-YFP* overexpressing strains and two *Ptnpf2* mutants after the shift to normal pH and low pH. A) rETR\textsubscript{max} values after 3, 24 and 48 hours from the shift to pH 8. B) rETR\textsubscript{max} values after 3, 24 and 48 hours from the shift to pH 7. Error bars represent standard deviations of three biological replicates. * indicates \( p \) < 0.05 and *** for \( p \) < 0.001.

No differences have been observed between strains in the PSII antenna functional cross-section (\( \sigma \) - Supplementary Figs. S5.2C-D), while a decrease in the relative reaction centres number has been observed in *Ptnpf2* knock-out mutants after 24 and 48 hours from the shift to low pH (Supplementary Figs. S5.2E-F), suggesting the decrease in photosynthetic efficiency could be partially due to structural reorganisation of PSII reaction centres after low pH shift.
Non-photochemical quenching (NPQ) is a photoprotective mechanism dissipating excess of light energy as heat. It relies on acidification of the lumenal pH, which affects pigment composition via the xanthophyll cycle (Bailleul et al., 2010). NPQ$_{\text{max}}$ was analysed among different strains and pH conditions. While after the shift to normal pH, no differences have been observed between strains (Figs. 5.18A, C, E), after 24 and 48 hours from the shift to low pH, $Ptnpf2$ knock-out strains show a small but significant increase of NPQ$_{\text{max}}$ compared to wild-type and overexpressing strains, suggesting a stress response (Figs. 5.18D, F). In particular, increases of 11-13% after 24 hours and of 8-10% after 48 hours from the shift to low pH have been observed in $Ptnpf2$ knock-out mutants compared to wild-type at the same time points and in the same pH condition. In general, NPQ response has been observed later in time (at 24 hours) compared with $F_{v}/F_{m}$ and ETR responses visible already 3 hours after the pH shift. This suggests acclimative responses of the $Ptnpf2$ knock-out strains to limit potential light stress by upregulating synthesis of molecular actors involved in NPQ, like xanthophyll cycle pigments and enzymes (Blommaert et al., 2021) and Light-Harvesting Complexes (Lhcx) proteins (Buck et al., 2019).
Figure 5.18. NPQ\textsubscript{max} values for \textit{P. tricornutum} wild-type, two \textit{PtNPF2-YFP} overexpressing strains and two \textit{Ptnpf2} mutants after the shift to normal pH and low pH. A) NPQ\textsubscript{max} values after 3, 24 and 48 hours from the shift to pH 8. B) NPQ\textsubscript{max} values after 3, 24 and 48 hours from the shift to pH 7. Error bars represent standard deviations of three biological replicates. * indicates $p < 0.05$ and *** for $p < 0.001$.

Considering E50\textsubscript{NPQ}, that is the irradiance level required to reach 50% of NPQ\textsubscript{max} and indicates the sensitivity to light intensity, \textit{Ptnpf2} knock-out mutants show significantly low E50\textsubscript{NPQ} compared to the other strains independently of time of measurements and the pH shift (Supplementary Figs. S5.4A-B). Moreover, overexpressing strains show higher E50\textsubscript{NPQ} than wild-type, in particular after 48 hours from the shift to low pH (Supplementary Fig. S5.4B). This indicates a NPQ induction more reactive to moderate light exposure in \textit{Ptnpf2} knock-out mutants compared to wild-type and overexpressing strains, confirming the stress response activated in \textit{Ptnpf2} strains in both pH conditions. On the other hand, no differences have been observed in nNPQ, which represents the Hill-
NPQ relaxation and $F'_v/F'_m$’ recovery were also studied, and measured parameters are reported in Supplementary Figure S5.1C. The $Ptnpf2$ knock-out mutants show a faster NPQ relaxation compared to wild-type, at all the time points of the shift experiment analysed, and after the shifts to both normal and low pH, as observed from their low $T_{50}$, which indicates the time needed to relax 50% NPQ (Fig. 5.19). On the opposite, PtNPF2 overexpressing strains show an increased $T_{50}$, suggesting a slower NPQ recovery, compared to wild-type, at all the time points analysed and after the shifts to both normal and low pH (Fig. 5.19). Speeding up the NPQ relaxation has been proven to be an effective strategy to increase photosynthetic performance, allowing cells to not wastefully dissipate light energy as heat when it is not saturating (Blommaert et al., 2021). NPQ dissipation is regulated by different factors, among which the lumenal pH and NADPH availability in the stroma, that could be directly or indirectly affected by PtNPF2 mutations.
Figure 5.19. NPQ half-time for relaxation ($T_{50}$) values for *P. tricornutum* wild-type, two *PtNPF2-YFP* overexpressing strains and two *Ptnpf2* mutants after the shift to normal pH and low pH. A) $T_{50}$ values after 3, 24 and 48 hours from the shift to pH 8. B) $T_{50}$ values after 3, 24 and 48 hours from the shift to pH 7. Error bars represent standard deviations of three biological replicates. * indicates $p < 0.05$; ** for $p < 0.01$ and *** for $p < 0.001$.

NPQ components were analysed: i) the energy-dependent quenching (qE), major contributor to NPQ occurring in the light harvesting complexes of PSII and rapidly responding to most short-term light stress, and ii) photo-inhibitory quenching (qI), associated with xanthophyll cycle reactions and including all components with slow relaxation kinetics (Lu et al., 2022). No large differences in NPQ components have been observed among strains as a function of time after the shift to both normal pH and low pH (Supplementary Fig. S5.5).

At the end of the NPQ relaxation, $F_v/F_m$ recovery was assessed again, so that percentage of PSII photosynthetic efficiency recovery was calculated. No significant differences have been observed in $F_v/F_m$ recovery (Supplementary Fig. S5.6).
So, photophysiology experiments show small but significant differences in $F_v/F_m$ and in ETR$_{max}$ in $Ptnp2$ knock-out mutants compared to wild-type after 3 hours from the shift to low pH; similar differences have been observed in NPQ$_{max}$ after 24 and 48 hours from the shift to low pH. Those phenotypes are consistent with a higher general stress in the $Ptnp2$ mutants at pH 7, but it is difficult to conclude whether such a small phenotype in photosynthesis can explain the 25% decrease in growth in the first 24 hours (Fig. 5.15F).

In addition to photosynthetic parameters, relative intracellular pH ($pH_i$) was measured. Recent studies showed that $P. tricornutum$ pH$_i$ decreases as the pH of the external medium (pH$_e$) decreases (Shi et al., 2019). This is consistent with the relative pH$_i$ trend in wild-type and overexpressing strains after the shift to low pH (Fig. 5.20B), while $Ptnp2$ knock-out mutants show relative pH$_i$s higher than wild-type after 24 and 48 hours, with values more similar to those measured at normal pH (Figs. 5.20A-B). This difference suggests an imbalance of the internal pH homeostasis of $Ptnp2$ knock-out mutants compared with the other strains, which could be responsible for the physiological differences observed.
**Figure 5.20.** Relative intracellular pH (pHᵢ) values for *P. tricornutum* wild-type, two *PtNPF2-YFP* overexpressing strains and two *Ptnpf2* mutants after 3, 24 and 48 hours from the shift to normal pH (pH 8 – A) and low pH (pH 7 – B). pHₑ represents the pH of the external medium. Error bars represent standard deviations of three biological replicates. * indicates $p < 0.05$; ** for $p < 0.01$ and *** for $p < 0.001$.

### 5.3.5. Dark to light shift

Previous gene expression studies showed a downregulation of *PtNPF2* during a prolonged dark growth (see paragraph 2.3.5 “Expression patterns of NPFs in *P. tricornutum*” in Chapter 2), indicating a possible light regulation. Moreover, photoprotection mechanisms, such as NPQ, are closely related to the H⁺ gradient across the thylakoid membranes (Blommaert et al., 2021): the higher lumenal pH, present during dark, decreases after light exposure (Höhner et al., 2016). This is a consequence of the stroma alkalinisation upon illumination, which pushes H⁺ from the stroma towards the lumen in order to reach the optimal pH condition to allow Calvin-Benson enzymes present on the stroma to work (Höhner et al., 2016). So, a shift experiment from dark and
light was performed, in order to probe these mechanisms and understand if there are any differences between strains in the photosynthetic and photoprotective capacities.

Maximum fluorescence, photosynthetic efficiency and electron transport rate increase after the shift to light reaching high values after 24 hours in all the strains (Figs. 5.21B, C and E). Maximum NPQ also increases upon illumination, reaching close to maximal values already after 3 hours (Fig. 5.21F). Besides expected differences in photosynthetic performances between dark and light conditions, no significant difference has been observed between strains. This suggests that, if PtNPF2 is involved in regulating H⁺ fluxes between chloroplast membranes upon dark to light shift, it is not the main player.
Figure 5.21. Photosynthetic parameters for *P. tricornutum* wild-type, two *PtNPF2-YFP* overexpressing strains and two *Ptnpf2* mutants before and after the shift from dark to light. A) Cell concentration, B) maximum fluorescence $F_{\text{max}}$, C) $F_v/F_m$, D) $\alpha$, E) $rE\text{TR}_{\text{max}}$ and F) $NPQ_{\text{max}}$ values before the shift (dark) and 3 and 24 hours after the shift to light. Error bars represent standard deviations of three biological replicates. * indicates $p < 0.05$.

5.3.6. Bicarbonate assay

A pilot experiment was performed in order to understand if PtNPF2 could be a HCO$_3^-$ transporter. In fact, carbonic anhydrases converting CO$_2$ into HCO$_3^-$, have been found in the PPC compartment, proposing a role of the PPC in the CCM mechanisms. So, oxygen evolution was measured after the addition of increasing amounts of HCO$_3^-$. Despite the lack of replicates, the results show a decrease of the maximum rate of photosynthesis in
both $Ptnpf2$ knock-out lines, compared to the wild-type strain (Fig. 5.22). This could reveal a possible alteration of the carbon fixation/acquisition pathways in the knock-out lines. Conversely, the overexpressing line 2OE 3 does not affect photosynthesis compared to the wild type, while the 2OE 2 strain has a lower maximum value of O$_2$ production (Fig. 5.22). The observations are in any case not still conclusive and must be confirmed with further experiments and replicates, which are in progress.

![Bicarbonate assay](image)

**Figure 5.22.** Rates of net photosynthesis as a function of HCO$_3^-$ concentration for f $P. tricornutum$ wild-type, two $PtNPF2$-YFP overexpressing strains and two $Ptnpf2$ mutants. Error bars are not present as these are preliminary results representing only one replicate.

### 5.3.7. Samples for RNA-seq

$P. tricornutum$ wild-type and $Ptnpf2$ knock-out mutant 2KO 1.15 were chosen to perform an RNA-seq analysis on the shift experiment from normal to low pH, in order to determine gene expression differences between strains and between the two pH conditions, and to provide more information about the PtNPF2 role. Because of the growth rate reduction and impaired photosynthetic parameters mainly observed after 24 hours from the shift to low pH, the shift experiment was repeated in bigger volumes, collecting samples just before the shift and after 24 hours from the shift to pH 7. Growth reduction in the $Ptnpf2$ knock-out mutant has been confirmed (Fig. 5.23A), so cells were collected, RNA was extracted (Fig. 5.23B) and quantified. Two samples (WT pH 8 samples 1 and 3) were re-extracted because of the low RNA amount of the initial samples (Fig. 5.23B).

Moreover, from the first PCR performed on RNA with intron spanning $PtNOA$ primers, a residual genomic DNA contamination was observed, as it is possible to see in lanes with
a 150 bp band (Fig. 5.23C). So, samples were treated with DNase I, another PCR with *PtNOA* primers was carried out and, once no genomic DNA contamination was confirmed by observing the agarose gel (Fig. 5.23D), these samples were sent to the sequencing company.

**Figure 5.23.** RNA samples preparation for RNA-seq analysis. A) *P. tricornutum* wild-type and *Ptnpf2* knock-out mutant 2KO 1.15 growing in flasks after 24 hours from the shift to low pH (pH 7). Labels indicate different strains. B) RNA extracted from cells and run on 1% agarose gel. Two samples (WT pH 8 samples 1 and 3) were re-extracted because of the low RNA amount of the
initial samples. M represents the 1Kb marker. C-D) PCR on PtNOA gene to check genomic DNA contamination in RNA samples before (C) and after DNase I treatment (D). 150 bp band indicates genomic DNA, 90 bp band indicates intron-less cDNA, here used as positive control of the PCR, and no band indicates absence of DNA in the preparation. M represents the 100 bp marker, as control for the gel quality.

5.4. Discussion

Ion balance is a delicate equilibrium regulating growth, fitness and ecological success of many organisms, however often poorly known. It is finely tuned by channels and transporters located on different cellular membranes, which allow exchanges with the external environment but also between different intracellular compartments and organelles.

Diatom NPFs is a family of transporters which diverges into two Clades (Santin et al., 2021b), of which Clade I diNPFs is phylogenetically and structurally close to bacteria POTs. Bacteria POTs are $\text{H}^+\text{-coupled peptide transporters}$ able to recognise amino acids, peptides, $\text{NO}_3^-$ and defence compounds, and drive the uptake of their specific substrates by utilising the electrochemical proton gradient across the membranes (Parker et al., 2017).

NPF2 of $P. \text{tricornutum}$ belongs to diNPFs Clade I (Santin et al., 2021b) and has been chosen as representative to study the possible role of this group of NPF transporters in diatoms.

From in silico prediction of the subcellular localisation (see paragraph 2.3.8 “Predicted subcellular localisation” of the Chapter 2), PtNPF2 was predicted to the plasma membrane, as all the other diNPF sequences belonging to the Clade I. Experimental validation of PtNPF2 subcellular localisation, performed in this Chapter through the generation and the observation of strains expressing PtNPF2 in frame with two different fluorescent tags, showed that PtNPF2 is actually localised in a small compartment close to the nucleus and between the two chloroplast lobes, which probably corresponds to the PPC compartment.

This hypothesis is supported from literature data, showing the same fluorescence pattern in correspondence of this small and still poorly known compartment associated with the chloroplast (Maier et al., 2015; Moog et al., 2011; Maier et al., 2022; Marter et al., 2020). Moreover, the consistency of C-terminal YFP and N-terminal GFP signals, observed in
independent transgenic lines, confirms the fluorescent signal is not an artifact but comes from a specific subcellular compartment.

PPC is a compartment located between the membranes of the diatom chloroplast, and it is surrounded by the PPM. Since diatom chloroplast owns four membranes as result of the secondary endosymbiosis, ions and molecules transport across them needs to be finely regulated. Both PPC and PPM have been proposed to have key roles in regulating the fluxes of molecules between cytosol and chloroplast, acting as communication link between cellular metabolism and photosynthetic activity (Maier et al., 2022). For example, some carbonic anhydrases or HCO$_3^-$ transporters have been predicted to be localised in this compartment, ensuring carbon and ions supply for C fixation (Tachibana et al., 2011).

However, BTS signal peptide, composed of a N-terminal signal sequence followed by a transit peptide-like region which addresses proteins to the chloroplast and which is often present in amino acidic sequences of previously PPC-localised enzymes (Maier et al., 2015), has not been found in the PtNPF2 sequence. The absence of the BTS is not crucial as it can be due to the weak prediction performance of available tools for transit peptide-like sequences of organisms that are of secondary endosymbiotic origin (Moog et al., 2011). Other sequences have been predicted to localised on PPC or PPM based on wide genomic analyses of conserved N-terminal extensions of protein sequences, difficult to unveil with currently available predictive tools (Moog et al., 2011).

The possible PPC localisation of PtNPF2 needs to take into account an important element: PtNPF2 is a transmembrane transporter, so needs membranes to correctly fold. In a recent study of Flori et al. (2016), a vesicular network (VN) was observed in the PPC. This VN, initially called periplastidial reticulum (Gibbs, 1979), did not extend around the whole chloroplast but was found restricted to particular locations, between the two lobes of the chloroplast and close to the nucleus, often present close to the chloroplast-nucleus isthmus (Fig. 5.24) (Flori et al., 2016). The VN could be involved in the transport of different molecules from the cytosol to the chloroplast, but its role as well as the proteins present on the membranes that shape it are still unknown.
Figure 5.24. Electron micrograph scanning of a *Phaeodactylum* PPC. A) Original micrograph and B) organelles and compartment reconstruction. The outer nuclear envelope is shown in blue and it is in continuity with the cERM. The VN fills the space between the PPM and oEM, shown in light green. C, chloroplast; M, mitochondrion; N, nucleus; Thyl, thylakoids. Image adapted from Flori et al. (2016).

The presence of the VN membranous structure in the PPC could explain the localisation of PtNPF2, a transmembrane transporter, in this small blob-like compartment, supporting the hypothesis of its possible role in substrate trafficking between the cytosol and the plastidial stroma.

Even if PtNPF2 substrates have not been identified yet, some hypotheses can be done based on results obtained through phenotypic characterisation of overexpressing strains and knock-out mutants generated with the CRISPR/Cas9 proteolistic method.

In fact, PtNPF2 showed a differential gene expression in response to different external pH, with a significant upregulation at low pH, which together with the presence of the ExxER structural motif responsible for H⁺-coupled bond, suggested a possible role of PtNPF2 in H⁺-transport and pH regulation (Santin et al., 2021b). Moreover, the knock-out of this gene caused a growth lag at low pH, which also showed an impaired relative intracellular pH and a higher NPQ capacity.

pH balance is very complex in aquatic photosynthetic organisms such as diatoms, as it occurs through the regulation of H⁺ and HCO₃⁻ fluxes, linked with C transport and fixation, but needs also to consider Na⁺, Cl⁻ and other ions present in the environment which flow across membranes and influence cell homeostasis. Exploration of these
mechanisms is still in its infancy, but some interesting suggestions can be taken from other organisms.

In fact, different mechanisms regulating ion balance and homeostasis have been studied in higher plants, recognising the role of many co-transporters, which can move anions and/or cations simultaneously through symport or antiport activities (Taylor et al., 2012; Léran et al., 2014). Among these, most NPF members function as active H⁺-symporters thanks to the ExxER motif (Newstead, 2015; Longo et al., 2018), while others have been found to work as proton-coupled H⁺/K⁺ antiporters, such as A. thaliana NPF7.3 (Li et al., 2017a). Generally, there are many plant examples of acquisition of an anion and a cation, such as the H⁺ and NO₃⁻ symport mediated by NPF6.3 in A. thaliana (Longo et al., 2018), or of the antiport of similarly charged ions, in order to satisfy organisms requests while maintaining charge balance (Feng et al., 2020).

Similarly, bacteria POTs, phylogenetically and structurally close to Clade I diNPFs, have the ability to recognise diverse ligands while retaining a strict requirement to couple transport to the electrochemical proton gradient (Parker et al., 2017). These transporters are fundamental not only for satisfying nutritional needs of the organism, but also to finely regulate the internal homeostasis of cells despite changing external environment.

Since in the pH shift experiments described in this Chapter, low culture pH has been reached by adding HCl to the medium, the main changing factor between normal and low pH condition is the H⁺ concentration. It follows that cells in which PtNPF2 has been knocked out showed a growth reduction probably because of increased H⁺ concentration, so supporting the hypothesis of a PtNPF2 role in H⁺ transport.

Results about relative intracellular pH measurements supported this hypothesis. In fact, in wild-type P. tricornutum, as well as other diatoms, intracellular pH trend follows extracellular pH values, so that pHᵢ decreases as environmental pH decreases (Shi et al., 2019). This has been confirmed by data reported in this Chapter. Moreover, Ptnpf2 knock-out mutants, showing growth reduction at low pH, in this condition presented a relative pHᵢ higher than wild-type and overexpressing strains, with values very close to those measured at normal pHₑ. This suggests an intracellular pH imbalance, which can be associated with the growth reduction and affected photobiology observed, possibly caused by an impaired H⁺ transport across intracellular compartments.
As previously said, many bacteria POTs and higher plants NPF transporters have been shown to transport H$^+$ or other ions together with different substrates, from small peptides to big molecules like phytohormones (Léran et al., 2014).

A possible PtNPF2 substrate co-transported together with H$^+$ could be HCO$_3^-$ . In fact, from preliminary results obtained through the pilot HCO$_3^-$ assay, an alteration of the carbon fixation/acquisition pathways in the *Ptnpf2* knock-out lines compared to wild-type was observed. Since beta-carbonic anhydrases, crucial components of the CCM and CO$_2$ fixation, have been found in *P. tricornutum* PPC (Moog et al., 2011; Maier et al., 2015; Tanaka et al., 2005; Tachibana et al., 2011), a putative PtNPF2 role in transporting HCO$_3^-$ could be hypothesised.

Nevertheless, other biochemical analyses will be needed to experimentally determine PtNPF2 substrates and test HCO$_3^-$ and H$^+$ transport hypothesis. These analyses will include heterologous expression of PtNPF2 in *Xenopus laevis*. In fact, *X. laevis* oocytes are a commonly used heterologous system for the expression and functional characterisation of membrane proteins, allowing to determine not only the molecules transported, but also the affinity and the transport kinetics. The oocytes have all the needed accumulated stores and wait, like a closed fully equipped factory, for a message of fertilisation (Miller and Zhou, 2000). *X. laevis* oocytes have been preferred in this study because this approach has been successfully employed to identify and characterise many NPFs in plants (Miller and Zhou, 2000). In particular, experiment will be performed under different conditions, such as different N sources or concentration, providing different substrates, and testing different pHs. Expression vectors to perform these analyses has already been generated (see paragraph 5.2.15 “Vector for *PtNPF2* heterologous expression in *Xenopus laevis*” in this Chapter), and experiments are now in progress.

As previously said, the PPC is a compartment localised between two of the four chloroplast membranes of diatoms, supporting chloroplast functioning and being responsible for molecules exchanges between cytosol and plastidial stroma (Maier et al., 2022; Flori et al., 2016). In this context, a missing protein in the PPC could affect the photosynthetic activity of the cell.

Results showed lower F$_v$/F$_m$ in *Ptnpf2* knock-out mutants compared with other strains, in the first hours after the shift to low pH. This transient phenotype could be directly or indirectly linked with the growth reduction observed in *Ptnpf2* knock-out mutants in that condition, confirming a functional role of PtNPF2 upon a low pH shift, although cause-
and-effect relationships can hardly be defined as of now. Moreover, the decrease in the relative reaction centres number observed in \textit{Ptnpf2} knock-out mutants after the shift to low pH suggested that the reduced photosynthetic efficiency could be partially due to structural reorganization of Photosystem II reaction centres after the low pH shift.

Also, electron transport rate was observed to be affected in the mutants, with a relative ETR$_{\text{max}}$ lower in the \textit{Ptnpf2} knock-out strains compared to wild-type during the first hours after the shift, and higher in overexpressing strains after 48 hours from the shift. These data were consistent between biological triplicates but, despite being significant, their modest amplitude is insufficient to completely explain the large growth reduction of \textit{Ptnpf2} knock-out strains (around 25% lower). Moreover, the Photosystem II efficiency in limited light (defined by $\alpha$ value) did not show any difference between strains, suggesting that in the range of light intensity close to the growth light used, PSII is equally effective between strains. Therefore, as far as can be investigated with dynamic fluorescence, the absence of PtNPF2 does not impede PSII functioning under moderate light intensity at low or normal pH, suggesting that other bioenergetic mechanisms are disturbed upon low pH shift in the \textit{Ptnpf2} knock-out and lead to their lower growth rates. These mechanisms could include alternative electron fluxes, carbon-fixation, or higher mitochondrial respiration.

Interestingly, the \textit{Ptnpf2} knock-out strains also showed a higher maximum NPQ compared to both wild-type and overexpressing strains, which appeared later in time compared to F$_{\text{v}}$/F$_{\text{m}}$ and ETR$_{\text{max}}$ differences, after the pH shift. Mechanisms activating NPQ are different from those involved in Photosystem II efficiency and electron transports, and include: i) short-term mechanisms which regulate NPQ kinetics or light response over a light curves, which mainly depend upon luminal pH and NADPH in the stroma, and ii) long-term acclimative processes prior to the light curved, depending on upregulation of xanthophyll cycle and Lhcx proteins which define the NPQ potential prior to measurements (Blommaert et al., 2021; Seydoux et al., 2022).

Moreover, a faster NPQ relaxation in \textit{Ptnpf2} knock-out strains compared to wild-type was observed (Fig. 5.19). While it is not pH shift dependent, as this difference was observed at both pH 8 and pH 7, it tells something about possible NPF2 function. NPQ relaxation kinetics is the sum of the rate of the Diadinoxanthin De-Epoxidase (DDE), which converts diadinoxanthin into diatoxanthin in the lumen and which is regulated by luminal pH, and the rate of the Diatoxanthin Epoxidase (DEP), which catalyses the
opposite reaction in the stroma and which depends upon an unknown regulation (probably stromal pH and/or redox state) with the need of NADPH as a co-factor (Blommaert et al., 2021). So, it is possible that in Ptnpf2 knock-out strains, either DDE is slower because of the imbalanced luminal pH or DEP is faster, or both.

Although it is very difficult to investigate such mechanisms because of different factors involved and because lumenal and stromal pH variations are very rapid and probes to measure them are very sensitive, a correlation between NPQ and presence/absence of PtNPF2 at saturating light intensity upon a low pH shift was observed. Lower F\textsubscript{v}/F\textsubscript{m} and ETR together higher NPQ are non-specific proxies of general stress. In particular, NPQ increase in Ptnpf2 knock-out strains suggests that PtNPF2 could be involved in mechanisms which regulate stress response and energy dissipation, with the higher NPQ capacity and the faster NPQ recovery of Ptnpf2 knock-out mutants suggested this response could be due more to aclimative changes.

To better investigate this response more analyses will be needed, such as the study of pigments involved in xanthophyll cycle, in particular diatoxanthin and diadinoxanthin (Blommaert et al., 2021). Moreover, an RNA-seq is now in progress to explore, among others, the expression levels of genes involved in short- and long-term NPQ mechanisms.

Other interesting information was given by dark to light shift experiment. In fact, in the dark the cytoplasmic and stromal pH are low, while upon illumination the stroma becomes alkaline, as a prerequisite for the full activation of pH-dependent Calvin Benson cycle enzymes, with the consequent H\textsuperscript{+}-pumping across the thylakoid membrane (Höhner et al., 2016). This implies a pH gradient between the cytosol and the stroma upon illumination, maintained by an active H\textsuperscript{+} export mechanism in the chloroplast membranes to compensate for passive H\textsuperscript{+} diffusion from cytosol to stroma (Höhner et al., 2016). However, no significant differences among strains were observed in photosynthetic performances after illumination, suggesting that PtNPF2, while involved in H\textsuperscript{+} transport across membranes, could not be the only player in these mechanisms.

So, in this Chapter new insights on PtNPF2 functional role have been described and new hypotheses have been proposed. PtNPF2 has been shown to putatively localise in the PPC, and to be probably involved in H\textsuperscript{+} transport across chloroplast membranes and pH regulation. Moreover, a possible role in H\textsubscript{CO}_3\textsuperscript{-} has been proposed. All these results move once again the focus of NPF transporters study from the interface with the external environment to the intracellular distribution of molecules and resources, touching new
fields of investigations which include intracellular ion balance and communication between different cell compartments.

However, other analyses are ongoing, in order to: i) confirm PPC subcellular localisation of PtNPF2, though the ongoing transformation of PtNPF2-YFP overexpressing strains with a plasmid containing a PPC-localised fusion protein which will allow co-localisation observations; ii) support NPQ results of Ptnpf2 knock-out mutants, through pigment analyses which will allow a more in depth understanding of xanthophyll cycle response; iii) study the gene expression profile of wild-type and Ptnpf2 knock-out strains in response to low pH, so determining molecular links between different metabolic pathways and key players, through a RNA-seq analysis already in progress; and iv) identifying PtNPF2 transported substrate, its affinity and kinetics, through heterologous expression of PtNPF2 in Xenopus laevis oocytes, for which plasmid has already been generated.
Chapter 6: General discussion
6.1. Thesis aim and main results

Diatoms are one of the most successful phytoplanktonic groups, with unique morphological, ecological and genomic features which enable them to cope with a strongly variable environment. Despite their ecological relevance, very little is known on the molecular mechanisms which allow diatoms to sense, transport and manage nutrients, among which N, mostly in the form of NO$_3^-$, represent one of the most important but often limiting ones (Sanz-Luque et al., 2015).

The general aim of my PhD thesis was to characterise a specific class of putative low-affinity NO$_3^-$ transporters, NPFs, in diatoms. The NPFs, together with high-affinity NO$_3^-$ transporters NRT2s, have been recently proposed to regulate NO$_3^-$ sensing and uptake (Rogato et al., 2015; Busseni et al., 2019), however none of them has been functionally characterised yet. While NRT2s are thought to be active at low NO$_3^-$ concentrations, similarly to those that can be found in the ocean, NPFs are thought to be active at higher NO$_3^-$ concentrations, so it is really unexpected that diatoms own a transporter family which works in conditions usually not present in their environment. Moreover, in many other organisms in which they have been widely studied and characterised, such as bacteria and higher plants, NPFs have been shown to recognise and transport a large number of substrates, besides NO$_3^-$ (Léran et al., 2014).

Among the main scientific achievements in this work, diNPFs were identified using available genomes but also meta-datasets like TARA Oceans and MMETSP. Moreover, their phylogeny and structural models were reconstructed and diNPFs expression patterns were studied. diNPFs diverge into two Clades, and most diatoms own two NPFs which belong to these two distinct Clades and whose expression is regulated by light and pH. Strongly informative were the results obtained from predicted subcellular localisation. This was one of the key points which moved subsequent questions and experiments: in fact, a significant percentage of sequences, all belonging to Clade II, was predicted to the tonoplast, an important N storage organelle in diatoms, shifting the point of view from the uptake to the reallocation of molecules (Santin et al., 2021b).

The first identification of diatom NPFs (diNPFs) together with the reconstruction of their evolution and structural models and the investigation of diNPFs expression patterns allowed to propose working hypotheses on diNPFs functioning, which were subsequently experimentally supported by the functional characterisation of NPFs in Phaeodactylum...
tricornutum. To address this last goal, also methodological challenges were addressed. In fact, a recent proteolistic protocol for the model diatom *P. tricornutum* (Serif et al., 2018), was here applied and optimised to generate *PtNPFs* knock-out mutants through the CRISPR/Cas9 system. Since that publication, this is the first work applying and improving the proteolistic protocol in diatoms, successfully generating knock-out mutants, while making the protocol more efficient and cost-effective (Russo et al., 2022).

So, *P. tricornutum* NPFs were overexpressed and knocked out to functionally characterise them, as representatives of the two main groups in which diNPFs are divided. The confirmed tonoplast localisation of PtNPF1, together with the delayed growth phenotype observed in knock-out mutants suggested a possible role of PtNPF1 in NO$_3^-$ reallocation between different subcellular compartments and in N fluxes regulation inside diatom cells. Moreover, I exploited strains overexpressing tonoplast localised PtNPF1 fused with a fluorescent tag, and a specific tonoplast tracker, as available tools to better investigate the vacuole morphology in *P. tricornutum*, with a focus on its behaviour in N starvation conditions. These observations, together with intracellular NO$_3^-$ content analyses, not only confirmed the key role of the vacuole in storage and reallocation of NO$_3^-$, but also highlighted its importance in other mechanisms, such as buoyancy and osmotic regulation.

On the other hand, PtNPF2 putative localisation on the periplastidial compartment and reduced growth phenotype observed in knock-out strains, linked to an imbalanced intracellular pH and to an increased photoprotection capacity, allowed to hypothesise that PtNPF2 could play a role in regulating H$^+$ transport between cytosol and chloroplast, and internal pH homeostasis. To approach the study of ion transports across chloroplast membranes is very challenging, because ions, in particular H$^+$ and HCO$_3^-$, are key players in regulating photosynthetic processes inside the chloroplast, and diatoms own four plastidial membranes as result secondary endosymbiosis (Flori et al., 2016). The fine regulation of ion transport across these membranes is still poorly understood, as well as the role of different compartments localised between them. So, identifying a putative H$^+$ transporter involved in ion exchange across chloroplast membranes represents an interesting starting point for broadening the comprehension of such mechanisms.

In general, although the substrates for none of the PtNPFs have been identified yet, results here described suggest that diatom NPFs evolved to transport and regulate intracellular nutrient and ion fluxes, rather than uptake them from the external environment. In this
way, NPFs could contribute to the ecological success of diatoms which sees them as the most competitive phytoplanktonic microorganisms in oligotrophic or highly variable sea waters.

6.2. Thesis discussion

This PhD thesis started from the ocean, through the identification of diatom low affinity NO$_3^-$ transporters, diNPFs, and the study of their global distribution, zooming into the intracellular compartments of diatom cells, through the functional characterisation of $P$. tricornutum NPFs mutants.

At the time I started my PhD, the study of N transporters in diatoms was quite limited: after a first characterisation of NH$_4^+$ transporters (AMT) in $Cylindrotheca$ fusiformis done by Hildebrand (2005), an interesting work reviewing diatom N transporters was written by Rogato et al. (2015), and a more recent meta-omics study of Busseni et al. (2019) gave more insight about diatom AMTs and high-affinity NO$_3^-$ transporters NRT2s. Taking inspiration from these previous works, I decided to focus on NPFs.

In Chapter 2, I reported the first study of diNPFs (Santin et al., 2021b). In detail, the nine available diatom genomes, together with TARA Oceans and MMETSP meta-datasets were interrogated to identify NPFs in diatoms. Interestingly, most diatoms were shown to own two NPFs, with some exceptions mainly due to multiple alleles (e.g. $F$. cylindrus) and polyploidy phenomena (e.g. $F$. solaris). The high conservation of NPFs in diatom genomes made me wonder what the role of these transporters was. The only exception was represented by Chaetoceros, one of the largest diatom genera in marine phytoplankton, in which we did not find NPFs, neither in transcriptomic analyses (Amato et al., 2017) nor in TARA or MMETSP datasets. NPFs absence may be due to the presence of alternative genes with conserved function or to different adaptations for N transport, maybe due to the strict interaction with N-fixing bacteria in the Chaetoceros phycosphere (Olofsson et al., 2019). More investigations will be needed to explain this exception (see below paragraph 6.3 “Future perspectives”).

From phylogenetic analyses, diNPFs were observed to diverge into two Clades, one closer to bacteria POTs and the other one closer to higher plant NPFs, and most diatoms were shown to own NPFs belonging to both Clades (Santin et al., 2021b). This phylogenetic subdivision was also supported by predicted structural models of diNPF proteins,
obtained from the available crystal structures of plant and bacteria NPFs, which allowed to confirm the structural similarity of Clade I diNPFs with bacteria POTs and of Clade II diNPFs with plant NPFs. Moreover, subcellular localisation of diNPFs was predicted, finding a significant percentage of sequences, around 25%, predicted to the tonoplast, all belonging to Clade II (Santin et al., 2021b), again highlighting a distinction between diNPFs Clade I and Clade II. These results gave interesting suggestions for further functional characterisation studies. I started hypothesising that NPFs belonging to the two different Clades could have different functions in diatoms, maybe localising on different membranes and/or transporting different substrates. In particular, the subcellular localisation predictions allowed to hypothesise that Clade II diNPFs may be involved in reallocation of molecules, such as NO$_3^-$, between different cell compartments like the vacuole, where NO$_3^-$ concentrations can be higher than in the external environment (Kamp et al., 2011). In fact, diatoms are able to survive in different stress conditions such as dark, N starvation or anoxia for long periods thanks to their ability to accumulate NO$_3^-$ in the vacuole up to 100 mM (Kamp et al., 2011). The hypothesis that the two diNPFs Clades should have different functions was crucial to prompt a shift in the research approach, moving from the study of the interface between cell and external environment to the intracellular environment.

Predicted structural models of diNPF proteins gave also the possibility to perform other analyses. They allowed to: i) identify structural differences among clades; ii) find key residues involved in the substrate bond, in particular the ExxER motif which couples H$^+$ binding to NO$_3^-$ transport, the tyrosine residues responsible for NO$_3^-$ bond and the glutamic acid residues putatively binding H$^+$; iii) hypothesise their functioning mechanism of alternate access, which involves two different conformations stabilised by salt bridges (see Chapter 2). These results, obtained through the comparison between diNPFs sequences with the large quantity of data coming from the structural and functional characterisation of plant and bacteria NPFs, confirm the high structural conservation of NPFs between organisms, and their peculiar functioning despite the non-specificity of the substrates transported (Corratgé-Faillie and Lacombe, 2017; Jørgensen et al., 2017; Léran et al., 2020; Newstead, 2017).

DiNPFs structural analysis could be further implemented. In particular, the key residues putatively involved in different bonds or needed for conformational changes represent a starting point for further investigation, aimed to point-mutate these residues and confirm
or disprove their role for the NPF functioning (see below paragraph 6.3 “Future perspectives”).

The diNPFs expression pattern was investigated through correlations between TARA data and environmental conditions, through a literature survey and through qPCR performed on P. tricornutum cultures. This integrated approach allowed to conclude that different N sources and different N concentrations do not affect diNPF genes expression. As NPFs were first annotated as NO$_3^-$ transporters, this lack of correlation gave important suggestions which can be translated into two different hypotheses: i) diNPFs are not sensitive to environmental N changes probably because they are not localised on the plasma membrane, supporting the predictions of subcellular localisation, according to which Clade II diNPFs could be involved in N reallocation between intracellular compartments rather than its uptake from the external environment, or ii) diNPFs are not sensitive to environmental N changes maybe because they do not transport exclusively NO$_3^-$, as suggested by previous studies on NPFs of other organisms, which showed that these transporters can carry a wide range of different substrates (Jørgensen et al., 2017; Léran et al., 2020; Newstead, 2017).

In line with the latter hypothesis, the observed effect of light and pH on diNPF genes expression, in particular affecting PtNPF2 gene expression, led to the idea that this transporter could be involved in H$^+$ transport. This is consistent with the conservation of the ExxER structural motif in all diNPFs analysed, shown to be involved in H$^+$ transport, supporting our hypothesis of a H$^+$-dependent transport for diNPFs (Longo et al., 2018; Newstead, 2017; Solcan et al., 2012). In fact, H$^+$ are widely recognised to be a cotransported substrate of many NPFs in other organisms, providing strong advantages for NPF transporters family that could accommodate chemically diverse ligands while retaining the conserved ability to couple transport to the H$^+$ electrochemical gradient (Parker et al., 2017). This idea was tested in Chapter 5, and integrated in the functional characterisation of PtNPF2, even if more investigations will be needed.

The first identification and integrated analysis on diNPFs, made by combining omics, phylogenetic, structural and expression analyses (Santin et al., 2021b) provided a robust ground for the further characterisation of P. tricornutum NPFs. In fact, P. tricornutum, a model diatom for which many genetic tools are available, owns exactly two NPFs, belonging to the two Clades, respectively and it represented a very good candidate for
further functional studies through the generation of different genetically engineered strains.

*P. tricornutum* overexpressing strains were produced for both *PtNPFs*, expressing *PtNPF1* and *PtNPF2*, respectively, in frame with a fluorescent tag which allowed to confirm some of the previous subcellular localisation predictions. The results produced for *PtNPF2* were particularly relevant since I could demonstrate localisation in a poorly characterised compartment, the periplastidial compartment (PCC) so far described only in a few publications (Flori et al., 2016; Maier et al., 2015; Moog et al., 2011; Tachibana et al., 2011; Tanaka et al., 2005), thereby increasing knowledge of *P. tricornutum* cell biology and providing new leads for understanding the role of the PPC. Then, as the best way to study the function of a gene is to switch it off and to study the consequent phenotypic changes, knock-out mutants of *P. tricornutum NPFs* were generated through the CRISPR/Cas9 system, using an innovative approach called proteolistic transformation (Serif et al., 2018).

*PtNPF1*, belonging to the plant-like diNPFs Clade II, was expressed in frame with two different fluorescent tags, allowing to observe and confirm that *PtNPF1* actually localises on the tonoplast. In fact, these observations support previous *in silico* prediction of diNPFs subcellular localisation, from which an interestingly high number of diNPF sequences were predicted to the tonoplast, all belonging to the Clade II (see Chapter 2 and Santin et al., 2021b). This is also consistent with the *PtNPF1* homology with low-affinity plant NPFs, active at high NO$_3^-$ concentrations, and the high NO$_3^-$ concentrations which can be reached in diatom vacuole, up to 100 mM (Kamp et al., 2011). Together, findings support the hypothesis that *PtNPF1* could be responsible for NO$_3^-$ reallocation from the vacuole towards cytosol.

At this point, questions emerged on the role of the vacuole in diatoms, an interesting and challenging field of investigation, for which there is still little knowledge. My results indicated that vacuole size, shape and number were not affected by N starvation, even if intracellular NO$_3^-$ shows strong variations: this supports the idea that changes in intracellular NO$_3^-$ could translate to a different texture or biochemical composition of the vacuole (McCarthy et al., 2017; Raven and Beardall, 2022), taking into account that this organelle also plays a critical role in osmotic regulation, predator defence and buoyancy control (Behrenfeld et al., 2021). The second reflection derives from the fact that diatoms prefer to uptake N from the environment, if available, without showing preferences for
NO$_3^-$ or NH$_4^+$, rather than consuming accumulated N in the vacuole, so maintaining intracellular N stored (see Chapter 4). This result, together with i) the energetic advantage of NH$_4^+$ uptake, ii) the NO$_3^-$ bioavailability in the ocean and iii) the capability of diatoms to store NO$_3^-$ in vacuole, suggests that probably an equilibrium needs to be reached through co-provision of different N sources (Glibert et al., 2016). This is a very interesting outcome, as it partially overcomes the idea that diatoms are NO$_3^-$ specialists (Glibert et al., 2016). Glibert and colleagues (2016) suggested the diatoms are indeed NO$_3^-$ specialists, while cyanobacteria and many chlorophytes and dinoflagellates, may be better adapted to use of NH$_4^+$. However, results here represent show that much has still to be learned about the physiological response that allows different organisms to cope with different nutrient supplies, especially reduced forms of N, and focus once again the attention on the complex N metabolisms of diatoms, for which many questions, in particular about N storage and reallocation, are still open.

A transient growth delay was observed in *Ptnpf1* knock-out mutants after N starvation and repletion, linked to a delay in N uptake and an altered gene regulation of the high-affinity transport system components (see Chapter 4). This result led to the hypothesis that PtNPF1 could be involved in the communication between different compartments such as the plasma membrane and the tonoplast, in particular acting as sensors of the physiological status of the cell and triggering a signalling pathway which results in the NO$_3^-$ fluxes regulation. In this idea, that still needs to be validated, PtNPF1 could be a key player in N fluxes management and N resources reallocation between different intracellular compartments, such as the vacuole, where NO$_3^-$ can reach very high concentrations (Kamp et al., 2011). PtNPF1 involvement in vacuolar NO$_3^-$ efflux towards cytosol could be of great importance to N reallocation and efficient utilisation by *P. tricornutum* cell, explaining how N taken up and stored in vacuoles can be subsequently mobilised. In this way, PtNPF1 could mediate NO$_3^-$ efflux similarly to AtNPF5.11, AtNPF5.12 and AtNPF5.16 which are responsible for vacuolar NO$_3^-$ release in *Arabidopsis* (He et al., 2017).

The preliminary functional characterisation of PtNPF1, leading to the interesting hypothesis of its involvement in intracellular N fluxes regulation, represents one of the main outcomes of my PhD thesis, but also a starting point for further investigation aimed at supporting or disproving this proposition, and at advancing in the gradual discovery of molecular mechanisms which regulates physiological responses to N (see below
paragraph 6.3 “Future perspectives”). Moreover, as many signals such as the recovery of the phenotype indicate that other players can be involved in these processes, identifying components of N signalling and understanding how they operate in different environmental and physiological conditions would be critical to shed light on the ecological basis of diatom success.

In parallel, the functional characterisation of PtNPF2, belonging to the bacteria-like diNPFs Clade I, was performed (see Chapter 5). The first interesting result came from the experimental validation of subcellular localisation prediction previously made: in fact, despite previous bioinformatic predictions localised PtNPF2 on the plasma membrane, this transporter was observed to localise on a small dot close to the nucleus which probably represents the PPC. This little studied compartment is placed between the four membranes of the diatom chloroplast, and hosts a vesicular network putatively involved in the transport of different molecules from the cytosol to the chloroplast (Flori et al., 2016). The PtNPF2 localisation on PPC suggests that PtNPF2 could have a key role in regulating the fluxes of molecules between cytosol and chloroplast, acting as communication link between cellular metabolism and photosynthetic activity (Maier et al., 2022).

For PtNPF2 knock-out mutants, following results described in Chapter 2 about differential PtNPF2 gene expression in response to light and pH, and the hypothesis of a H\(^+\)-dependent transport, different shift experiments were carried out. Again, I observed a transient phenotype, which led to the idea that molecular mechanisms investigated are more complex than what expected, involving more than one actor which contributes to shape physiological responses. Results obtained supported previous hypotheses about PtNPF2 involvement in H\(^+\) transport, and inspire future directions in the PtNPF2 study.

In detail, an impaired growth was observed in PtNPF2 knock-out mutants after the shift from normal to low pH, together with imbalanced relative intracellular pH (see Chapter 5). These main results, in agreement with the putative PPC localisation of PtNPF2, support the idea that this transporter could be really involved in H\(^+\) transport across chloroplast membranes and in internal pH homeostasis. This was supported also by the high photoprotective capacity observed in PtNPF2 knock-out mutants, photoprotective capacity which has been shown to be strictly connected with lumenal pH regulation (Blommaert et al., 2021).
The exchanges between cytosol and chloroplast across the four plastidial membranes of diatoms have been receiving more interest in the last years (Liu et al., 2022; Seydoux et al., 2022), but are still poorly understood, as well as the transporters which regulate those exchanges. Given the PtNPF2 localisation and the impaired phenotype of Ptnpf2 knock-out mutants in response to increasing H⁺ concentration, the contribution of PtNPF2 to H⁺ transport between cytosol and chloroplast can be proposed, but still needs to be confirmed.

Particular attention should be dedicated to the putative substrate of PtNPF2: gene expression analyses and phenotypic characterisation of Ptnpf2 knock-out strains support the involvement of the protein in H⁺ transport. Different mechanisms regulating ion balance and homeostasis have been studied in higher plants, recognising many NPFs as co-transporters, able move anions and/or cations simultaneously through symport or antiport activities (Taylor et al., 2012; Léran et al., 2014). Among these, most NPF members are active H⁺-symporters, thanks to the ExxER motif (Newstead, 2015; Longo et al., 2018) such as the NPF6.3 in A. thaliana (Longo et al., 2018), or H⁺-antiporters, such as A. thaliana NPF7.3 (Li et al., 2017a). Similarly, bacteria POTs, phylogenetically and structurally close to Clade I diNPFs, have the ability to recognise diverse ligands in a H⁺-dependent manner (Parker et al., 2017). In this context, PtNPF2 transport of H⁺ needs to be confirmed, as well as possible other substrates still need to be identified (see below paragraph 6.3 “Future perspectives”). A first candidate is represented by HCO₃⁻: in fact, from preliminary results described in Chapter 5, an alteration of the carbon fixation/acquisition pathways was observed in the Ptnpf2 knock-out lines.

All these results move once again the focus of NPF transporters study from the interface with the external environment to the intracellular distribution of molecules and resources, touching new fields of investigations which include intracellular ion balance and communication between different cell compartments.

Therefore, this is the first study aimed to characterise NPF transporter family in diatoms. Coming back to the main question of this project “Why do diatoms own low-affinity NO₃⁻ transporters?”, it is now possible to say that diatom NPFs are not involved in NO₃⁻ uptake from the external environment, but they are rather involved in the regulation of intracellular fluxes of molecules. Other diatom transporters have been shown to localise on organelle membranes: for example, Pᵢ transporters are present on the tonoplast, involved in Pᵢ storage, or in particular endomembranes, whose role has to be defined yet.
(Dell’Aquila et al., 2020), while many transporters of N compounds have been predicted to mitochondrion and chloroplast envelopes, to allow N fluxes between different intracellular compartments (Smith et al., 2019). Si transporters have been shown to localise on specific vesicles necessary for biosilification processes (Shrestha and Hildebrand, 2015). This distribution and coordination of diatom membrane transport is necessary to maintain optimal metabolism, growth and cell division in response to fluctuating resource availability in the environment, but also to cell physiological needs (Brownlee et al., 2022). Coordination of transport pathways relies on the cross-talk between sensing and signalling of one or different substrates, to affect the whole regulation of transport and metabolic pathways of those substrates (Brownlee et al., 2022). Therefore, these results shift diatom study to a more complex and “self-reflective” point of view, as diatoms responses are no longer seen as mainly determined by environmental conditions, but rather as a combination of extracellular and intracellular signals which dynamically integrate both environmental conditions and physiological status.

Finally, the main achievements of this thesis were not only scientific but also methodological.

In fact, as regards to methodological challenges addressed, previously mentioned knock-out mutants of \textit{P. tricornutum} NPFs were generated through the CRISPR/Cas9 system, using an innovative approach called proteolistic transformation (Serif et al., 2018). This method had two main advantages, compared to previous techniques: an antibiotic-free selection and the reduction of off-target mutations. However, there were still some limitations, in fact the high amount of Cas9 needed translated to high costs for transformation experiments. The application and optimisation of this protocol was carefully described in Chapter 3, and published in Russo et al. (2022).

The achievements of this optimisation process were the increase of the gRNAs:Cas9 molar ratio, increasing INDELs frequency, and the decrease up to five times the amount of the nuclease Cas9 needed for each transformation experiment. These improvements allowed to further reduce the possibility to obtain off-target mutations, due to the reduced permanence of Cas9 nuclease into cells, and to make this technique more cost-effective. Since the first paper on proteolistics application in diatoms was published (Serif et al., 2018), this is the solely work in which the same method was successfully applied to
generate knock-out mutants, but also optimised, making it more manageable, accessible and reproducible.

In addition, the pipeline to screen mutants was carefully described, in order to select clones with biallelic mutations, which caused a frameshift or a stop codon insertion resulting in truncated and inactive proteins. Following this pipeline, two Ptnpf1 knock-out strains and two Ptnpf2 knock-out strains were selected to be used for further functional characterisation. Moreover, double Ptnpf1-2 mutants were generated and, even if no strains with biallelic mutation in both genes have been isolated yet, more transformation experiments will be done in the future to achieve this goal.

Therefore, the technological work presented in Chapter 3 represented a key step for the process of improving scientific knowledge about diNPFs (after the first identification of diNPFs in the Chapter 2, towards the functional characterisation of PtNPFs, presented in Chapters 4 and 5 and still ongoing), but also an important achievement from a methodological point of view, after the first publication of proteolistics in diatom from Serif et al. (2018), towards more complex applications aimed to knock-out gene families (Hao et al., 2021) or to study native regulatory elements (Nam et al., 2022).

6.3. Future perspectives

Even if these results provide new insights into the possible role of NPFs in diatoms, further investigation will be needed to support proposed hypotheses. Following the outline of the thesis, results obtained often helped to add a new piece to the complex characterisation of NPFs in diatoms, but at the same time entailed new questions.

First of all, after the identification of diNPFs I observed that no NPF sequences were retrieved for *Chaetoceros*. Why? It could be explained by different adaptations for N transport, which could benefit from the strict interaction with N-fixing bacteria. However, in the next future, the project “Most Abundant Diatom Genus in the World’s Ocean” (https://jgi.doe.gov/csp-2020-most-abundant-diatom-denus-world-ocean/) will provide genomic and transcriptomic data for many *Chaetoceros* species and, when they are available, it will be possible to obtain confirmation of the absence of NPFs in the genus and more information about the alternative strategies in place to fulfil the same role. This can be an interesting starting point for new studies aimed to better understand how biologic interactions can shape cell responses to environmental changes and how they can
drive evolution. Today, this is an emerging field of study, with challenges in finding adequate model organisms to manipulate in laboratory and/or in developing new tools to investigate this kind of interactions, in laboratory and in the field.

A new approach, which could help the research on nutrient fluxes occurring between different interacting organisms, and how they drive evolution, could be represented by fluxomics. Recently Smith et al. (2019) used fluxomics and flux balance analyses to examine short-term shifts in N utilisation in *P. tricornutum* cells, obtaining a systems-level understanding of assimilation and intracellular distribution of N. Among the new tools, allowing to perform flux analyses at cellular or intracellular levels, there is NanoSIMS, able to map the spatial distribution of trace elements and isotopes (such as $^{15}$N) with a nano- to micrometre scale resolution. It was first applied by Lechene et al. (2007) that used NanoSIMS to quantify N$_2$ fixation by individual bacteria, and is spreading in studies on C, N and O assimilation, fixation, transport and subcellular storage bodies (Musat et al., 2012).

This tool can be particularly useful also in the context of PtNPF1 functional characterisation. In fact, the hypothesis of its role in sensing physiological N conditions, to then regulate intracellular N fluxes, could obtain further insight from a single-cell technique. The same experiment in N starvation and repletion could be repeated (see Chapter 4) in combination with isotope labelling and analysed with NanoSIMS, as starting point for new and more complex experiments. In particular, this tool can map N uptake but also N movements between different intracellular compartments, helping to explain still unclear processes such as luxury uptake, N vacuolar storage and its reallocations. Moreover, NanoSIMS could definitely confirm if PtNPF1 is a NO$_3^-$ transporter, by comparing intracellular fluxes between wild-type, overexpressing and knock-out mutants previously obtained.

In fact, the questions “What molecules do NPFs transport? What are the possible substrates? Are they really low-affinity NO$_3^-$ transporter?” still need some clear answers. I suggested PtNPF1 could transport NO$_3^-$ from the vacuole to the cytosol, because of the phenotype observed after N starvation and repletion, and because of the high NO$_3^-$ amount stored in the vacuole. Moreover, I proposed H$^+$ as main ion transported by PtNPF2, because of the phenotype observed after the shift to low pH, and because of the intracellular pH imbalance. It is important to note that most bacteria and plant NPFs
transport $H^+$ together with other molecules. However, no univocal answer has been given on PtNPFs substrates.

In this context, heterologous expression of PtNPFs in *Xenopus laevis* oocytes provides a powerful system for the expression and characterisation of plant membrane proteins (Miller and Zhou, 2000). In fact, many different plant membrane proteins and transporters have already been characterised using this expression system, so that it and the related biochemical assays should definitely enlighten the PtNPFs substrates nature and their affinity, defining also the transport kinetics. These data integrated with the ones obtained from PtNPFs functional characterisation, will give a complete overview of the NPFs functioning in diatoms. I already generated plasmids for the heterologous expression of PtNPF1 and PtNPF2 respectively. Analyses are ongoing in the laboratory of Dr. Benoit Lacombe at the Institut des Sciences des Plantes in Montpellier (France). These analyses will confirm if PtNPF1 finally transports $NO_3^-$ and its affinity for the substrate, and if PtNPF2 transports also $H^+$, $HCO_3^-$ and/or other molecules.

Identifying PtNPFs substrates would be a fundamental goal in understating their functioning and their role in diatoms. But PtNPFs are just representatives of the two diNPFs Clades identified (see Chapter 2). So, how is it possible to report PtNPFs findings to all other members of the same Clade? In Chapter 2, I identified structural differences and key residues putatively involved in different bonds or putatively functional for conformational changes. Demonstrating that the substrate is bound by specific key residues that are conserved among one or both Clades, could help to broaden to all diNPFs the assumptions experimentally made on NPFs of the model *P. tricornutum*.

In fact, an implemented approach to study not only the substrate transported by PtNPFs, but also the amino acids involved in binding those molecules, could be the site-directed mutagenesis. This tool is widely used in other organisms such as higher plants to alter key residues involved in NO$_3^-$ or $H^+$ bonds: for example, it was used in maize to demonstrate that the His356 of NPFs forms an electrostatic interaction with the NO$_3^-$ molecule (Wen et al., 2017). This goal can be achieved by heterologous expression or by using the CRISPR/Cas9 approach to generate a single-amino-acid-substitution (Ma et al., 2017). This approach has been rarely used to study microalgal protein functioning through the complementation of deficient yeast mutants (Ouyang et al., 2016), but it has been never applied to study endogenous proteins directly in microalgae, such as diatoms. The conserved amino acids identified in diNPFs (see Chapter 2), and corresponding to
key residues involved in substrate bond of bacteria or plant NPFs, can be exploited through this approach. Among these, tyrosines binding NO$_3^-$, glutamic acids binding H$^+$, residues involved in salt-bridges and a threonine putatively involved in phosphorylation processes. Directly mutating these residues could be an interesting challenge, both from a technical and a scientific point of view, to experimentally support hypothesised structural features of diNPFs and confirm their alternate-access functioning mechanism.

Another new and powerful tool which could implement the findings in key residues, structural models and protein functioning is represented by AlphaFold (https://alphafold.ebi.ac.uk/). It is an Artificial Intelligence system developed by DeepMind and EMBL’s European Bioinformatics Institute, able to predicts protein’s 3D structures starting from amino acid sequences with competitive accuracy. In the next future, it will be surely interesting to compare and improve obtained results with AlphaFold analyses, by adding information about protein interactions (Johansson-Åkhe and Wallner, 2022) and predict structural and phenotypic effects of single mutations (McBrude et al., 2023). This last promising tool, now being implemented with AlphaFold2, will allow to predict the effect of point mutation in key residues, so to expect a particular structural change or phenotype, which could drive future functional characterisation and experimental analyses.

If this last questions, and last approaches described, can improve the understanding of all diNPFs, there are questions that are addressed to individual PtNPFs, and that require special attention. Two of these questions are more specific, both regarding PtNPF2, and the experiments needed to provide clear answers are already ongoing, hoping for more confirmation in the short term.

The first question is about the confirmation of the periplastidial compartment localisation of PtNPF2. In fact, while PtNPF1 subcellular localisation to the tonoplast has been confirmed through a specific tonoplast tracker, for PtNPF2 the subcellular localisation to the periplastidial compartment was hypothesised thanks to a literature survey but has not been confirmed yet. Dr. Uwe Maier at the Philipps University of Marburg (Germany) kindly provided me with a plasmid containing a fusion protein which has been shown to experimentally localise to the periplastidial compartment (Marter et al., 2020). I am transforming $PtNPF2$-$YFP$ overexpressing strains with this plasmid, in order to subsequently perform co-localisation observations. To achieve this goal, an alternative strategy could be the immunogold labelling of diatom sections with an anti-GFP antibody,
that allows the localisation of proteins at the electron microscopy resolution level and the quantification of signals.

The second question aims to clarify the relationships between PtNPF2 and other metabolic pathways. To answer to this question, during the low pH shift experiment, where the Pt

Ptnpf2

knock-out phenotype was observed, samples were collected for pigment analyses and RNA extraction. Pigment analyses will give more information about Pt

Ptnpf2

knock-out photoprotective capacity, investigating the contribution of xanthophyll cycle in the higher NPQ phenotype. On the other hand, RNA has been extracted from samples collected and an RNA-seq is ongoing, in order to explore global gene expression modifications in response to low pH and PtNPF2 knock-out, and to define relationships between PtNPF2 and other metabolic pathways.

Other hints for reflection come from the PtNPF1 outcome of Chapter 4. My PhD thesis focuses on low-affinity NPFs, but results obtained from PtNPF1 functional characterisation showed that it is not the only key player in N transport between different intracellular compartments. Three high-affinity NO

3-

transporters (NRT2s) have been predicted to the tonoplast, and one of them has been shown to be differentially regulated in Pt

Ptnpf1

knock-out mutants (see Chapter 4). This suggests other investigations on the role of the vacuole and on intracellular N fluxes will need to take into account also NRT2 family. So how to better investigate the link between PtNPF1 and NRT2s in the context of N metabolism and transport? Different approaches can drive further research aimed to answer, for example more in depth investigations on gene expression through transcriptomics. An alternative approach could be the generation and functional characterisation of Pt

NRT2s

knock-out mutants (Tan, 2020) and even more challenging and fascinating the simultaneous knock-out of PtNPF1 and selected NRT2s, to better study the contribution of these actors in the complex N metabolism, and maybe could allow to obtain more stable phenotypes.

Another interesting hypothesis which needs to be better investigated regards PtNPF1 involvement in sensing and signalling mechanisms. Again, the transient phenotype in response to environmental N variations and the slow regulation of N-related genes expression in knock-out mutants suggested a cross-talk between different compartments, which need to communicate their relative N availability and induce metabolic re-modelling, to better combine cell requirements with intra- and extracellular N supply.
In this context, sensing and signalling studies are still challenging, even if in the last years more research groups are focusing on mechanisms which allow communication between different microorganisms or between different compartments of the same cell. However, there are some limitations. First of all, the molecular machinery underpinning nutrient sensing and signalling remains largely uncharacterised in diatoms. Moreover, how extra- and intracellular stimuli impact genome expression is still largely unknown, because it requires the integrated action of signalling components and genetic and epigenetic factors, that we are just starting to discover in these microalgae (Jaubert et al., 2022). For this reason, experiments aimed to verify sensing capacity of specific proteins are often complicated and poorly conclusive. Moreover, PtNPFs have been shown to localise to intracellular membranes, rather than to the plasma membrane, which is the interface with the external environment: this means that there is no direct interaction between the protein of interest and the external environment, which can be manipulated, making the investigation more difficult to understand.

The final general and challenging question that I want to propose, as concluding sentence of my thesis and at the same time starting point for more future studies, is about the ecological relevance of the cross-talk between different intracellular compartments and pathways. Biology is developing faster and faster, not only in terms of knowledge and tools, but also in term of multidisciplinary way of thinking, widening more and more the vision. For example, I started my PhD project on diatom responses to changes in the availability of an environmental abiotic factor and, on the way, I became ever more conscious of the strong links between different environmental factor, between abiotic and biotic factors, between extra- and intracellular environment. These links are the keys to understand how diatoms combine (but also generate) stimuli to optimise their whole physiology and metabolism, and how this influence and their ecological role in the ocean. This could be an interesting starting point for future diatom studies which are becoming increasingly oriented to microorganisms sensing and communication.


Do, P.T., et al., 2019. Demonstration of highly efficient dual gRNA CRISPR/Cas9 editing of the homeologous GmFAD2–1A and GmFAD2–1B genes to yield a high oleic, low linoleic and α-linolenic acid phenotype in soybean. BMC Plant Biology, 19(1), p.311.


Appendices
Appendix A. Supplementary materials relative to Chapter 2

Supplementary Table S2.1

Complete set of sequences aligned and used to build the phylogenetic tree presented in Figure 2.1 and Supplementary Figure S2.1. Bold names represent clades shown as collapsed branches in Figure 2.1 tree.

BACTERIA POT (outgroup):
- Bac_Ecol_YbgH_Escherichia_coli_pdb_4Q65
- Bac_Ecol_Yjdl_Escherichia_coli
- Bac_Ecol_YdgR_Escherichia_coli
- Bac_Ecol_YhiP_Escherichia_coli
- Bac_Sone_PepT2_Shiwanella_oneidensis_pdb_4LEP
- Bac_Gkau_PepT1_Geobacillus_kaudophilus_pdb_4KV
- Bac_Sthe_PepT1_Streptococcus_thermophiles_pdb_4APS_5MMT
- Bac_Yent_PepT1_Yersinia_enterococila_pdb_4W6V

BOLIDOPHYCEAE:
- Bolidomonas_pacifica_Ochrophyta_CAMPEP_0118633580
- Bolidomonas_pacifica_Ochrophyta_CAMPEP_0182473638
- Bolidomonas_sp_Ochrophyta_CAMPEP_0197568754

PELAGOPHYCEAE:
- Aureococcus_anophagefferens_P_CAMPEP_0168944714
- Aureococcus_anophagefferens_P_CAMPEP_0169002006
- Aureococcus_anophagefferens_P_CAMPEP_0169039532
- Pelagococcus_subviridis_P_CAMPEP_0195701936
- Pelagococcus_subviridis_P_CAMPEP_0195762098
- Pelagococcus_subviridis_P_CAMPEP_0195808332
- Pelagococcus_subviridis_P_CAMPEP_0195747808

BACILLARIOPHYCEAE CLADE I:
- Amphiprora_paludosa_B_b_CAMPEP_0172475720
- Entomoneis_sp_B_b_CAMPEP_0198155306
- fs15278
- fs26459
- sro22590
- 'P.tricornutum_ID47218__'
- pt47218
- Grammatophora_oceanica_B_f_CAMPEP_0194044684
- sro193560
- sro262280
- Asterionellopsis_glacialis_B_f_CAMPEP_0184897570
- Asterionellopsis_glacialis_B_f_CAMPEP_0195308738
- Asterionellopsis_glacialis_B_f_CAMPEP_0195319058
- Asterionellopsis_glacialis_B_f_CAMPEP_0197148292
- Asterionellopsis_glacialis_B_f_CAMPEP_0195288138
- Asterionellopsis_glacialis_B_f_CAMPEP_0197142392
- Cyclophora_tenuis_B_f_CAMPEP_0116552302
- Extubocellulus_spinner_B_m_CAMPEP_0178631424
- Extubocellulus_spinner_B_m_CAMPEP_0178661016
- Extubocellulus_spinner_B_m_CAMPEP_0178721124
- Odontella_B_m_CAMPEP_0113526884
- Minutocellus_polymorphus_B_m_CAMPEP_0181056214
<table>
<thead>
<tr>
<th>Species Name</th>
<th>Genbank Accession</th>
<th>Genbank Accession</th>
<th>Genbank Accession</th>
<th>Genbank Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minutocellus polymorphus B_m_CAMPEP_0181068454</td>
<td>Minutocellus polymorphus B_m_CAMPEP_0197729970</td>
<td>Proboscia_alata B_c_CAMPEP_0194399688</td>
<td>Proboscia_alata B_c_CAMPEP_0194423398</td>
<td>Proboscia_inermis B_c_CAMPEP_0171306902</td>
</tr>
<tr>
<td>Stephanopyxis turris B_c_CAMPEP_0195536914</td>
<td>Cylindrotheca closterium B_b_CAMPEP_0113603326</td>
<td>Cylindrotheca closterium B_b_CAMPEP_0113634624</td>
<td>Cylindrotheca closterium B_b_CAMPEP_0113617026</td>
<td>Attheya septentrioralis B_c_CAMPEP_0198287936</td>
</tr>
<tr>
<td>Dactyliosolen fragilissimus B_c_CAMPEP_0184858680</td>
<td>Leptocylindrus B_c_CAMPEP_0196816786</td>
<td>Leptocylindrus B_c_CAMPEP_0196822330</td>
<td>Leptocylindrus danicus B_c_CAMPEP_0116007456</td>
<td>Leptocylindrus danicus B_c_CAMPEP_0116068160</td>
</tr>
<tr>
<td>Cylindrotheca closterium B_b_CAMPEP_0113685264</td>
<td>Pseudo-nitzschia pungens B_b_CAMPEP_0172359006</td>
<td>Pseudo-nitzschia pungens B_b_CAMPEP_0172389482</td>
<td>Pseudo-nitzschia australis B_b_CAMPEP_0168167802</td>
<td>Pseudo-nitzschia australis B_b_CAMPEP_0168299120</td>
</tr>
<tr>
<td>Pseudo-nitzschia australis B_b_CAMPEP_0168253056</td>
<td>Pseudo-nitzschia arenysensis B_b_CAMPEP_0116124116</td>
<td>Pseudo-nitzschia delicatissima B_b_CAMPEP_0116088998</td>
<td>Pseudo-nitzschia delicatissima B_b_CAMPEP_0116109638</td>
<td>Nitzschia punctata B_b_CAMPEP_0178734884</td>
</tr>
<tr>
<td>Nitzschia sp B_b_CAMPEP_0113462588</td>
<td>Nitzschia sp B_b_CAMPEP_0113512952</td>
<td>Staurosira_constricta sp B_f_CAMPEP_0119573420</td>
<td>Bac_HAO78866.1_MFS_transporter__Verrucomicrobia_subdivision_3_bacterium_</td>
<td>Bac_PZO64660.1_MFS_transporter__Pseudomonas_swinnonensis_</td>
</tr>
<tr>
<td>Bac_WP_056131761.1_MFS_transporter__Lysobacter_sp.<em>Root494</em></td>
<td>Bac_PZR11506.1_MFS_transporter__Archangium_gephyra_</td>
<td>Bac_WP_141734295.1_MFS_transporter__Oligoflexus_tunisiensis_</td>
<td>Bac_Wc1_B_JX407531.1_MIFS_transporter__Dyella_</td>
<td>Bac_Sone_PepT1_Sphingomonas_campestris_pdb_6EI3</td>
</tr>
<tr>
<td>Bac_Wc1_B_JX407531.1_MIFS_transporter__Dyella_</td>
<td>Bac_WP_026633197.1_MULTI SPECIES_MFS_transporter__Dyella_</td>
<td>Bac_Sone_PepT1_Shewanella_oneidensis_pdb_2XUT</td>
<td>Bac_WP_040099119.1_MIFS_transporter__Oleiagrimonas_soli_</td>
<td></td>
</tr>
<tr>
<td>Bac_WP_068060530.1_MFS_transporter__Rheinheimera_sp.<em>SA_3</em></td>
<td>Bac_RYD14046.1_MFS_transporter__Xanthomonadaceae_bacterium_</td>
<td>Bac_WP_019539070.1_MFS_transporter__Rheinheimera_sp.<em>SA_3</em></td>
<td>Bac_WP_040099119.1_MIFS_transporter__Oleiagrimonas_soli_</td>
<td></td>
</tr>
</tbody>
</table>

**Bacteria MFS/POT:**

- Bac_HAO78866.1_MFS_transporter__Verrucomicrobia_subdivision_3_bacterium_
- Bac_PZO64660.1_MFS_transporter__Pseudomonas_swinnonensis_
- Bac_WP_056131761.1_MFS_transporter__Lysobacter_sp._Root494_
- Bac_PZR11506.1_MFS_transporter__Archangium_gephyra_
- Bac_WP_141734295.1_MFS_transporter__Oligoflexus_tunisiensis_
- Bac_Wc1_B_JX407531.1_MULTI SPECIES_MFSTransporter__Dyella_
- Bac_Sone_PepT1_Shewanella_oneidensis_pdb_2XUT
- Bac_WP_068060530.1_MFS_transporter__Rheinheimera_sp._SA_3_
- Bac_RYD14046.1_MFS_transporter__Xanthomonadaceae_bacterium_
- Bac_WP_040099119.1_MIFS_transporter__Oleiagrimonas_soli_

**P.PATENS NPF 8.5-8.6:**

Ppat_NPF8.5_was_Ppat_NPF7.1_P_patens_Pp1s9_211V6.1_Pp3c20_6850V3.1' 'Ppat_NPF8.6_was_Ppat_NPF7.2_P_patens_Pp1s69_25V6_Pp3c14_23410V3.1'

**C.ELEGANS OPTs/H.SAPIENS PepTs:**

291
Hsap_SCF15A1_PepT1_Homo_sapiens
Hsap_SCF15A2_PepT2_Homo_sapiens
OPT-1_Caenorhabditis_elegans
OPT-2_Caenorhabditis_elegans
OPT-3_Caenorhabditis_elegans

Fungi PTRs:
Fgra_PTR2A_Fusarium_graminearum
PTR2a_Saccharomyces_cerevisiae
PTR2_Saccharomyces_cerevisiae
Fgra_PTR2B_Fusarium_graminearum
Fgra_PTR2C_Fusarium_graminearum
Fgra_PTR2D_Fusarium_graminearum

Bacillariophyceae Clade II:
sro337320
Aulacoseira_subarctica_B_c_CAMPEP_0172418398
Aulacoseira_subarctica_B_c_CAMPEP_0172434738
Skeletonema_marinoi_B_c_CAMPEP_0115985824
Skeletonema_marinoi_B_c_CAMPEP_0180842762
Skeletonema_marinoi_B_c_CAMPEP_0197230968
Skeletonema_marinoi_B_c_CAMPEP_0180858880
sm2929
Skeletonema_marinoi_B_c_CAMPEP_0115997118
Skeletonema_dohrnii_B_c_CAMPEP_0206678184
Skeletonema_menzelii_B_c_CAMPEP_0183671166
fc147192
'jgi_Frac1_147192_gw1.15.276.1'
MATOU-v1_116232058_2_
MATOU-v1_110765158_3_
MATOU-v1_112767895_3_
MATOU-v1_112767894_3_
MATOU-v1_97032942_6_
Nitzschia_sp_B_b_CAMPEP_0113461804
Fragilariopsis_kerguelensis_B_b_CAMPEP_0170764078
Fragilariopsis_kerguelensis_B_b_CAMPEP_0171013162
Fragilariopsis_kerguelensis_B_b_CAMPEP_0196056204
Fragilariopsis_kerguelensis_B_b_CAMPEP_0170841070
Fragilariopsis_kerguelensis_B_b_CAMPEP_0196008772
Fragilariopsis_kerguelensis_B_b_CAMPEP_0196122576
MATOU-v1_113685789_1_
fc200740
'jgi_Frac1_200740_e_gw1.56.35.1'
MATOU-v1_116136333_5_
Fragilariopsis_kerguelensis_B_b_CAMPEP_0196095434
MATOU-v1_62720_5_
to32021
MATOU-v1_64811423_2_
MATOU-v1_51669342_6_
MATOU-v1_64649983_2_
MATOU-v1_80364487_2_
Aulacoseira_subarctica_B_c_CAMPEP_01724330662
cc26601
Cyclotella_meneghiniana_B_c_CAMPEP_0172267716
Thalassiosira_weissflogii_B_c_CAMPEP_0203126740
Thalassiosira_weissflogii_B_c_CAMPEP_0203392310
Thalassiosira_weissflogii_B_c_CAMPEP_0203590012
Thalassiosira_weissflogii_B_c_CAMPEP_0203425700
Thalassiosira_weissflogii_B_c_CAMPEP_0203616374
Thalassiosira_weissflogii_B_c_CAMPEP_0203495852
Nitzschia_punctata_B_b_CAMPEP_0178822004
Cylindrotheca_closterium_B_b_CAMPEP_0113627218
Thalassionema_nitzschioides_B_f_CAMPEP_0194250086
Thalassiothrix_antarctica_B_f_CAMPEP_0194161678
Thalassiothrix_antarctica_B_f_CAMPEP_0194196856
"F._cylindrusID_171976_"
c171976
MATOU-v1_116768328_6_
MATOU-v1_112321054_3_
MATOU-v1_97467216_1_
MATOU-v1_81344760_3_
MATOU-v1_90559390_5_
MATOU-v1_11381763_2_
MATOU-v1_13814119_2_
MATOU-v1_61380803_2_
Fragilariopsis_kerguelensis_B_b_CAMPEP_0195978708
MATOU-v1_113669673_2_
Pseudo-nitzschia_australis_B_b_CAMPEP_0168203384
Pseudo-nitzschia_australis_B_b_CAMPEP_0168298824
Pseudo-nitzschia_pungens_B_b_CAMPEP_0172379760
Pseudo-nitzschia_pungens_B_b_CAMPEP_0172411790
Nitzschia_punctata_B_b_CAMPEP_0178745106
Nitzschia_punctata_B_b_CAMPEP_0178832876
Nitzschia_punctata_B_b_CAMPEP_0178794896
Nitzschia_punctata_B_b_CAMPEP_0178813952
Nitzschia_punctata_B_b_CAMPEP_0178764652
Nitzschia_punctata_B_b_CAMPEP_0178857212
Nitzschia_sp_B_b_CAMPEP_0113453074
Nitzschia_sp_B_b_CAMPEP_0113521326
fs20187
fs25493
Amphora_coffeaeformis_B_b_CAMPEP_0170706364
sro2800
c0035672
Detonula_conservacea_B_c_CAMPEP_0172303608
Thalassiosira_antarctica_B_c_CAMPEP_0201993928
MATOU-v1_98255502_1_
Thalassiosira_sp_B_c_CAMPEP_0181095040
Thalassiosira_sp_B_c_CAMPEP_0181133398
Thalassiosira_weissflogii_B_c_CAMPEP_0203293664
Thalassiosira_weissflogii_B_c_CAMPEP_0203476302
Thalassiosira_weissflogii_B_c_CAMPEP_0203540154
Thalassiosira_weissflogii_B_c_CAMPEP_0201803906
Thalassiosira_weissflogii_B_c_CAMPEP_0203595428
Thalassiosira_weissflogii_B_c_CAMPEP_0203410916
Thalassiosira_weissflogii_B_c_CAMPEP_0203342628
Thalassiosira_weissflogii_B_c_CAMPEP_0203384942
Thalassiosira_sp_B_c_CAMPEP_0172343830
T_pseudonana_ID_7452
tp269333
Cyclotella_meneghiniana_B_c_CAMPEP_0172290586
Staurosira_constricta_sp_B_f_CAMPEP_0119555258
Staurosira_constricta_sp_B_f_CAMPEP_0119559140
c136520
MATOU-v1_113975913_2_
fc256377
'jgi_Fracy1_256377_fgenesh2_pg.75___68'
MATOU-v1_112320600_5_
MATOU-v1_1353625_2_
Pseudo-nitzschia_arenysensis_B_b_CAMPEP_0116121222
MATOU-v1_97116754_2_
Pseudo-nitzschia_delicatissima_B_b_CAMPEP_0116106712
Pseudo-nitzschia_delicatissima_B_b_CAMPEP_0116114374
MATOU-v1_19564035_2_
MATOU-v1_81873041_4_
Pseudo-nitzschia_pungens_B_b_CAMPEP_0172374276
Pseudo-nitzschia_pungens_B_b_CAMPEP_0172405090
Pseudo-nitzschia_australis_B_b_CAMPEP_0168180732
Pseudo-nitzschia_australis_B_b_CAMPEP_0168191072
Pseudo-nitzschia_australis_B_b_CAMPEP_0168283962
Pseudo-nitzschia_australis_B_b_CAMPEP_0168299840
Pseudo-nitzschia_australis_B_b_CAMPEP_0168248576
PM0007930
'PSNMU-V1.4_AUG-EV-PASAV3_0007930.1_'
pma7930
Nitzschia_sp_B_b_CAMPEP_0113455026
Nitzschia_sp_B_b_CAMPEP_0113504808
Attheya_septentrionalis_B_c_CAMPEP_0198306490
Cylindrotheca_closterium_B_b_CAMPEP_0113628188
Chaetoceros_B_c_CAMPEP_0176481300
Nitzschia_punctata_B_b_CAMPEP_0178748050
Nitzschia_punctata_B_b_CAMPEP_0178755690
'P.tricornutum__ID47148'
pt47148
Amphiprora_paludosa_B_b_CAMPEP_0172452160
Amphora_coffeaeformis_B_b_CAMPEP_0170703300

DINOPHYCEAE CLADE I:
Brandtodinium_nutriculum_Dinophyceae_CAMPEP_0198503000
Brandtodinium_nutriculum_Dinophyceae_CAMPEP_0198606146
Scrippsiiella_trochoidea_Dinophyceae_CAMPEP_0111570034
Kryptoperidinium_foliaceum_Dinophyceae_CAMPEP_0176049366
Kryptoperidinium_foliaceum_Dinophyceae_CAMPEP_0176206372
Durinskia_baltica_Dinophyceae_CAMPEP_0170225160
Pelagodinium_beii_Dinophyceae_CAMPEP_0197711576
'Symbiodinium_sp._Dinophyceae_CAMPEP_0169791830'
'Symbiodinium_sp._Dinophyceae_CAMPEP_0169847968'
'Symbiodinium_sp._Dinophyceae_CAMPEP_0169900436'
'Symbiodinium_sp._Dinophyceae_CAMPEP_0181548348'
'Symbiodinium_sp._Dinophyceae_CAMPEP_0181640646'
'Symbiodinium_sp._Dinophyceae_CAMPEP_0181712314'
'Symbiodinium_sp._Dinophyceae_CAMPEP_0201353458'

295
'Symbiodinium_sp._Dinophyceae_CAMPEP_0201370046'
'Symbiodinium_sp._Dinophyceae_CAMPEP_0181508290'
'Symbiodinium_sp._Dinophyceae_CAMPEP_0196911640'
'Symbiodinium_sp._Dinophyceae_CAMPEP_0181424080'
'Prorocentrum_minimum_Dinophyceae_CAMPEP_0205229678'
'Symbiodinium_sp._Dinophyceae_CAMPEP_0181437590'
'Scrippsiella_trochoidea_Dinophyceae_CAMPEP_0115186758'
'Scrippsiella_trochoidea_Dinophyceae_CAMPEP_0115664016'
'Scrippsiella_trochoidea_Dinophyceae_CAMPEP_0115394368'
'Symbiodinium_sp._Dinophyceae_CAMPEP_0181553564'
'Scrippsiella_trochoidea_Dinophyceae_CAMPEP_0115375790'
'Scrippsiella_hangoei_Dinophyceae_CAMPEP_0177394426'
'Scrippsiella_hangoei_Dinophyceae_CAMPEP_0204139724'
'Scrippsiella_trochoidea_Dinophyceae_CAMPEP_0115442672'
'Scrippsiella_trochoidea_Dinophyceae_CAMPEP_0115684800'
'Symbiodinium_sp._Dinophyceae_CAMPEP_0201393446'
'Symbiodinium_sp._Dinophyceae_CAMPEP_0202682524'
'Noctiluca_scintillans_Dinophyceae_CAMPEP_0194489542'
'Oxyrrhis_marina_Dinophyceae_CAMPEP_0205035916'
'Oxyrrhis_marina_Dinophyceae_CAMPEP_0204290210'
'Cryptothecodinium_cohnii_Dinophyceae_CAMPEP_0206586774'
'Scrippsiella_hangoei_Dinophyceae_CAMPEP_0177449446'
'Symbiodinium_sp._Dinophyceae_CAMPEP_0181565352'
'Prorocentrum_minimum_Dinophyceae_CAMPEP_0205648780'

HAPTOPHYTA:

Chrysochromulina_rotalis_haptophya_CAMPEP_0115848524
Prymnesium_parvum__Haptophya__CAMPEP_0113253836
Prymnesium_parvum__Haptophya__CAMPEP_0113293786
Prymnesium_parvum__Haptophya__CAMPEP_0113296870

DINOPHYCEAE CLADE II:

Durinskia_baltica_Dinophyceae_CAMPEP_0170374928
Durinskia_baltica_Dinophyceae_CAMPEP_0170396858
Durinskia_baltica_Dinophyceae_CAMPEP_0170440254
Durinskia_baltica_Dinophyceae_CAMPEP_0170395090
Kryptoperidinium_foliaceum_Dinophyceae_CAMPEP_0176184250

VIRIDIPLANTAE (Chlorophyceae NRT1/2 + Bryophyta NPFs Clade II + Spermatophyta):

'AT1G12110AtNPF6.3'
'AT2G26690AtNPF6.2'
'AT3G21670AtNPF6.4'
'Ppat_NPF6.4_P_patens_Pp1s298_45V6.1_Pp3c4_21340V3.1'
'Ppat_NPF6.6_P_patens_Pp1s133_58V6__Pp3c13_4730V3.2'
'AT1G27040AtNPF4.5'
'AT1G69850AtNPF4.6'
'AT5G62730AtNPF4.7'
'AT3G25260AtNPF4.1'
'AT3G25280AtNPF4.2'
'Ppat_NPF4.1_P_patens_Pp1s434_30V6_Pp3c17_16320V3.5'
'AT1G33440AtNPF4.4'
'AT1G59740AtNPF4.3'
'Ppat_NPF4.2_was_6_1_P_patens_Pp1s143_133V6.1_Pp3c4_7710V3'
'Ppat_NPF4.3_was_6_2_P_patens_Pp1s72_96V6.1_Pp3c5_1870V3.1_'
'AT1G18880AtNPF2.9'
'AT3G47960AtNPF2.10'
'AT5G62680AtNPF2.11'
'AT1G27080AtNPF2.12'
'AT1G69870AtNPF2.13'

296
'AT1G69860AtNPF2.14'
'AT3G45650AtNPF2.7'
'AT3G45660AtNPF2.6'
'AT3G45680AtNPF2.3'
'AT3G45700AtNPF2.4'
'AT3G45710AtNPF2.5'
'AT1G82190AtNPF1.2'
'AT3G16180AtNPF1.1'
'AT5G11570AtNPF1.3'
'AT1G68570AtNPF3.1'
'
'AT3G01350AtNPF5.9'
'AT5G14940AtNPF5.8'
'AT2G37900AtNPF5.6'
'
'AT1G32450AtNPF7.3'
'AT4G21680AtNPF7.2'
'AT5G19640AtNPF7.1'
'
'Csub_NRT1_Coccomyxa_subellipsoidea
'Cvar_NRT1_Chlorella_variabilis
'Czof_NRT2_Chromochloris_zofingiensis
'Crei_NRT1_Chlamydomonas_reinhardtii
'Cvar_NRT1_Volvox_carteri
'Dsal_NRT1_Dunaliella_salina
'Csub_NRT2_Coccomyxa_subellipsoidea
'Cvar_NRT1_Chlorella_variabilis

297
Supplementary Table S2.2

Diatom NPFs data coming from *TARA* Oceans dataset, including taxonomic assignation of MATOU sequences and corresponding Clades.

<table>
<thead>
<tr>
<th>MATOU ID</th>
<th>Taxon (Genus/Species)</th>
<th>Clade</th>
</tr>
</thead>
<tbody>
<tr>
<td>MATOU-v1_112774712_1_</td>
<td><em>Fragilariopsis kerguelensis</em></td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_91046335_3_</td>
<td><em>Fragilariopsis kerguelensis</em></td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_113688670_4_</td>
<td><em>Fragilariopsis kerguelensis</em></td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_81873041_4_</td>
<td><em>Pseudo-nitzschia delicatissima</em></td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_19564035_2_</td>
<td><em>Pseudo-nitzschia delicatissima</em></td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_97116754_2_</td>
<td><em>Pseudo-nitzschia delicatissima</em></td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_1353625_2_</td>
<td><em>Pseudo-nitzschia arenysensis</em></td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_112320600_5_</td>
<td><em>Pseudo-nitzschia</em></td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_113975913_2_</td>
<td><em>Fragilariopsis cylindrus</em></td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_98255502_1</td>
<td><em>Fragilariopsis cylindrus</em></td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_113669673_2</td>
<td><em>Fragilariopsis cylindrus</em></td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_13814119_2</td>
<td><em>Fragilariopsis cylindrus</em></td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_61380803_2</td>
<td><em>Fragilariopsis cylindrus</em></td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_11381763_2</td>
<td><em>Fragilariopsis cylindrus</em></td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_90559390_5_</td>
<td><em>Fragilariopsis cylindrus</em></td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_81344760_3</td>
<td><em>Fragilariopsis cylindrus</em></td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_98962776_1</td>
<td><em>Fragilariopsis cylindrus</em></td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_97467216_1_</td>
<td><em>Fragilariopsis cylindrus</em></td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_112321054_3_</td>
<td><em>Fragilariopsis cylindrus</em></td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_116768328_6_</td>
<td><em>Fragilariopsis cylindrus</em></td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_116457729_6_</td>
<td><em>Thalassiothrix antarctica</em></td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_98382034_4</td>
<td><em>Thalassiosira sp.</em></td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_98142622_4</td>
<td>Thalassiosira sp.</td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_97464270_5</td>
<td>Thalassiosira</td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_97464269_5</td>
<td>Thalassiosira</td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_72263465_3</td>
<td>Thalassiosira</td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_44301921_2</td>
<td>Detonula confervacea</td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_80364487_2</td>
<td>Unknown</td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_64649983_2</td>
<td>Unknown</td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_51669342_6</td>
<td>Unknown</td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_64811423_2</td>
<td>Unknown</td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_62720_5</td>
<td>Thalassiosira oceanica</td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_116136333_5</td>
<td>Fragilariopsis cylindrus</td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_113685789_1</td>
<td>Fragilariopsis kerguelensis</td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_97032942_6</td>
<td>Fragilariopsis</td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_112767894_3</td>
<td>Fragilariopsis</td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_112767895_3</td>
<td>Fragilariopsis</td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_110765158_3</td>
<td>Fragilariopsis</td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_116232058_2</td>
<td>Fragilariopsis</td>
<td>Diatom Clade II</td>
</tr>
<tr>
<td>MATOU-v1_19401141_4</td>
<td>Pseudo-nitzschia</td>
<td>Diatom Clade I</td>
</tr>
<tr>
<td>MATOU-v1_113685264_2</td>
<td>Fragilariopsis kerguelensis</td>
<td>Diatom Clade I</td>
</tr>
</tbody>
</table>

Supplementary Figure S2.1

Correlation, by means of Pearson’s $r$, of diNPFs DNA (metagenomics) and mRNA (metatranscriptomics) occurrences in the TARA Oceans dataset with (A) NO$_2^-$ and NO$_3^-$ and (B) PO$_4^{3-}$ concentrations. Scatter plots represent correlations for the following classes and depths: 5-20 µm, DCM; 5-20 µm, SUR; 20-180 µm, DCM and 20-180 µm, SUR.
Supplementary Figure S2.2

Correlation, by means of Pearson’s $r$, of diNPFs DNA (metagenomics) and mRNA (metatranscriptomics) occurrences in the TARA Oceans dataset with (A) Si and (B) Fe concentrations. Scatter plots represent correlations for the following classes: 5-20 µm, DCM; 5-20 µm, SUR; 20-180 µm, DCM and 20-180 µm, SURB. 20-180 µm, SUR; C. 5-20 µm, DCM; D. 20-180 µm, DCM.
Supplementary Figure S2.3

Correlation, by means of Pearson’s $r$, of diNPFs DNA (metagenomics) and mRNA (metatranscriptomics) occurrences in the TARA Oceans dataset with (A) SST (horizontal Sea Surface Temperature gradient), (B) MLE (Maximum Lyapunov Exponent) and (C) Chl $a$ (Chlorophyll $a$ concentration). Scatter plots represent correlations for the following classes: 5-20 $\mu$m, DCM; 5-20 $\mu$m, SUR; 20-180 $\mu$m, DCM and 20-180 $\mu$m, SURB. 20-180 $\mu$m, SUR; C. 5-20 $\mu$m, DCM; D. 20-180 $\mu$m, DCM.
Appendix B. Supplementary materials relative to Chapter 3

Supplementary File S3.1

PtAPT, PtNPF1 and PtNPF2 wild-type genomic sequences, including underlined introns.

>PtAPT (ID: Phatr3_J68349)
ATGACGACACAAAGGAGTTGGCAGCATGCTAGTGAGAGCAAAGTACGCCCAGGACGGCGACGAAGCCAAGGAAATTGCCGAGTACCTCCCTACCTCACTGCTGGTTTTTCTGTACGAGCCC
AAAGTCTTTCAAAAGATTGTAGACGTCTTTGTCGATCGGTACCGGGAAATCGGTGTCGACGTCATTGCGGGGTAAGTCTG
AAGGAAAGTGGGTTGGCCAGCGTGTCTGTGTGACACTTTCTCACCCTCTAGTGCTTGCACCACGACCATGCCTGGTAC

>PtNPF1 (ID: Phatr3_J47148)
ATGACGACACAAAGGAGTTGGCAGCATGCTAGTGAGAGCAAAGTACGCCCAGGACGGCGACGAAGCCAAGGAAATTGCCGAGTACCTCCCTACCTCACTGCTGGTTTTTCTGTACGAGCCC
AAAGTCTTTCAAAAGATTGTAGACGTCTTTGTCGATCGGTACCGGGAAATCGGTGTCGACGTCATTGCGGGGTAAGTCTG
AAGGAAAGTGGGTTGGCCAGCGTGTCTGTGTGACACTTTCTCACCCTCTAGTGCTTGCACCACGACCATGCCTGGTAC

>PtNPF2 (ID: Phatr3_J47218)
ATGAGTAGGAGAAGTTCTCGTGCACAAAGTAGGCATCTCGTTCCCCATTCTCCATCCGATACCGATACTTTCGCCCATAAAGACCGGCCATCATGCCCAGAGGCAACGGCACCACCCATAGTGTTCGGGAAAAGGATAAACTTTTGACGCCC
GTCATTTGGTGCATTCTCGTAACGGAAACGGGCGAGCGCTTCGCGTACTTTGGCTTTCGGGCAATTCTGGTGCTGTATTT
TTCGTCTCTCTGGAATATTCCGAATCTCAAGCGATTGCCTTTTTCGCGTATACTACCTGTTTGGCTTATTTGTCACCA
ATTGCTGGCGCGCTATTGGCCGACGGACACTTGGGACGTTACCAAACAATCCTGTGGTTCGGCCTCGTTTACGTGATT
GGCTTGTCCATTCTGACCTTCGCAGCAGCGGCATCCGAAGATGTAGATCTCGCGTATCGCCGAACGTTAACCTTCGTGG

>PtNPF2 (ID: Phatr3_J47218)
ATGAGTAGGAGAAGTTCTCGTGCACAAAGTAGGCATCTCGTTCCCCATTCTCCATCCGATACCGATACTTTCGCCCATAAAGACCGGCCATCATGCCCAGAGGCAACGGCACCACCCATAGTGTTCGGGAAAAGGATAAACTTTTGACGCCC
GTCATTTGGTGCATTCTCGTAACGGAAACGGGCGAGCGCTTCGCGTACTTTGGCTTTCGGGCAATTCTGGTGCTGTATTT
TTCGTCTCTCTGGAATATTCCGAATCTCAAGCGATTGCCTTTTTCGCGTATACTACCTGTTTGGCTTATTTGTCACCA
ATTGCTGGCGCGCTATTGGCCGACGGACACTTGGGACGTTACCAAACAATCCTGTGGTTCGGCCTCGTTTACGTGATT
GGCTTGTCCATTCTGACCTTCGCAGCAGCGGCATCCGAAGATGTAGATCTCGCGTATCGCCGAACGTTAACCTTCGTGG

303
Supplementary Table S3.1

Screening of positive colonies growing on selective medium after proteolistic shots. All the experiment performed to generate \textit{Ptapt}, \textit{Ptapt-Ptnpf1} and \textit{Ptapt-Ptnpf2} single mutants and \textit{Ptapt-Ptnpf1-Ptnpf2} double mutants are summarised.
<table>
<thead>
<tr>
<th>strain</th>
<th>genes knocked out</th>
<th>shot number</th>
<th>total positive colonies on selective medium</th>
<th>PCR screened clones</th>
<th>clones showing INDELs on gel</th>
<th>sequenced clones</th>
<th>clones with biallelic not functional mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>PtAPT</td>
<td>5</td>
<td>&gt;100</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>WT</td>
<td>PtAPT-PtNPF1</td>
<td>5</td>
<td>&gt;200</td>
<td>36</td>
<td>20</td>
<td>11</td>
<td>2 (heterozygous)</td>
</tr>
<tr>
<td>WT</td>
<td>PtAPT-PtNPF2</td>
<td>5</td>
<td>&gt;200</td>
<td>32</td>
<td>13</td>
<td>9</td>
<td>2 (heterozygous)</td>
</tr>
<tr>
<td>WT</td>
<td>PtAPT-PtNPF1-PtNPF2</td>
<td>10</td>
<td>&gt;200</td>
<td>77</td>
<td>17 (one PtNPF) 3</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Ptnpf1</td>
<td>PtUMPS-PtNPF2</td>
<td>10</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ptnpf2</td>
<td>PtUMPS-PtNPF1</td>
<td>5</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Supplementary File S3.2
Sequence alignment of PtAPT wild-type and knock-out mutant clone 4.
PtAPT_wt
ptAPT_clone4

1 ATGACGACAACCAACGGAAGTTCGGCAGCTGTGCTAGTGAGCGACAAGTA
1 ATGACGACAACCAACGGAAGTTCGGCAGCTGTGCTAGTGAGCGACAAGTA
**************************************************

50
50

PtAPT_wt
ptAPT_clone4

51 CGCCCAGGACGGCGACGAAGCCAAGGAAATTGCCGAGTACCTGCCCTACT
51 CGCCCAGGACGGCGACGAAGCCAAGGAAATTGCCGAGTACCTGCCCTACT
**************************************************

100
100

PtAPT_wt
ptAPT_clone4

101 TTCCCTTCAAAGGCATTCCACGCTTCTACGATATTGGTGGTTTTCTGTAC
101 TTCCCTTC-----------------------------------------********

150
108

PtAPT_wt
ptAPT_clone4

151 GAGCCCAAAGTCTTTCAAAAGATTGTAGACGTCTTTGTCGATCGGTACCG
109 --------------------------------------------------

200
108

PtAPT_wt
ptAPT_clone4

201 GGAAATCGGTGTCGACGTCATTGCGGGGTAAGTCAAGGAAAGTGGGTTGG
109 --------------------------------------------------

250
108

PtAPT_wt
ptAPT_clone4

251 CCAGCGTGTCTGTGTGACACTTTCTCACCCTCTAGTGCTTGCACCACGAC
109 --------------------------------------------------

300
108

PtAPT_wt
ptAPT_clone4

301 CACTCCTGTAGTTTGGACGCACGTGGGTTTGTCCTGGGTCCACCAATTGC
109 --------------------------------------TCCACCAATTGC
************

350
120

PtAPT_wt
ptAPT_clone4

351 CCTGGCTTTGAATAAACCCTTTATAATGATGCGGAAAAAGGGAAAAATGC
121 CCTGGCTTTGAATAAACCCTTTATAATGATGCGGAAAAAGGGAAAAATGC
**************************************************

400
170

PtAPT_wt
ptAPT_clone4

401 CCAACAGCGTATCCTCGGACGACTACACGACGGAATACGGTAACCGACAA
171 CCAACAGCGTATCCTCGGACGACTACACGACGGAATACGGTAACCGACAA
**************************************************

450
220

PtAPT_wt
ptAPT_clone4

451 GGCTTGACGGTGCAAAAAGACAAGATTCAAAAGAACCACCGGGTTCTGAT
221 GGCTTGACGGTGCAAAAAGACAAGATTCAAAAGAACCACCGGGTTCTGAT
**************************************************

500
270

PtAPT_wt
ptAPT_clone4

501 CATTGACGATTTGGTAGCCACGGGCGGAACCCTTGGGTCGGCCGTTAGTT
271 CATTGACGATTTGGTAGCCACGGGCGGAACCCTTGGGTCGGCCGTTAGTT
**************************************************

550
320

PtAPT_wt
ptAPT_clone4

551 TGGTCAAAATGCTCGGTGGCGTGGTCGTGGAGTGTGCCTGTGTGGTAGAG
321 TGGTCAAAATGCTCGGTGGCGTGGTCGTGGAGTGTGCCTGTGTGGTAGAG
**************************************************

600
370

PtAPT_wt
ptAPT_clone4

601 CTAAAAATGTTCATTGATCCCCCGGAGGAATCCGGTTTGCCCAGTCGGAA
371 CTAAAAATGTTCATTGATCCCCCGGAGGAATCCGGTTTGCCCAGTCGGAA
**************************************************

650
420

PtAPT_wt
ptAPT_clone4

651 AAAGCTCTTTGCCGAACTGGGTCACCACGATGTCGGTGTGTGGGGTTTGA
421 AAAGCTCTTTGCCGAACTGGGTCACCACGATGTCGGTGTGTGGGGTTTGA
**************************************************

700
470

PtAPT_wt
ptAPT_clone4

701 TCTCGGAAGACATCTTGACCAACAAAGCGGATCTCCCGCAGCACTACGTG
471 TCTCGGAAGACATCTTGACCAACAAAGCGGATCTCCCGCAGCACTACGTG
**************************************************

750
520

PtAPT_wt
ptAPT_clone4

751 GATGACGGCGAAGAACACTGA
521 GATGACGGCGAAGAACACTGA
*********************

306

771
541


**PtNPF1 and PtNPF2** wild-type and knock-out mutants used for subsequent phenotypic analyses. Schemes of biallelic mutations occurred on both alleles, with yellow gRNAs, grey insertions and dotted deletions. Sequence alignment of wild-type PtNPFs and mutated ones.

---

**Supplementary File S3.3**

PtNPF1 and PtNPF2 wild-type and knock-out mutants used for subsequent phenotypic analyses. Schemes of biallelic mutations occurred on both alleles, with yellow gRNAs, grey insertions and dotted deletions. Sequence alignment of wild-type PtNPFs and mutated ones.

---

**PtNPF1 and PtNPF2** wild-type and knock-out mutants used for subsequent phenotypic analyses. Schemes of biallelic mutations occurred on both alleles, with yellow gRNAs, grey insertions and dotted deletions. Sequence alignment of wild-type PtNPFs and mutated ones.

---

**PtNPF1 and PtNPF2** wild-type and knock-out mutants used for subsequent phenotypic analyses. Schemes of biallelic mutations occurred on both alleles, with yellow gRNAs, grey insertions and dotted deletions. Sequence alignment of wild-type PtNPFs and mutated ones.
PtNPF1_wt | ATACCGGGCGTTTTGGGAGATACTTTTAACCGTCGCGCCCTTTCAATTGGTCTCCTCGTC | 900
ptnpf1_2.8_allele1 | ATACCGGGCGTTTTGGGAGATACTTTTAACCGTCGCGCCCTTTCAATTGGTCTCCTCGTC | 900
ptnpf1_2.8_allele2 | ATACCGGGCGTTTTGGGAGATACTTTTAACCGTCGCGCCCTTTCAATTGGTCTCCTCGTC | 900

PtNPF1_wt | TTGTGGCCCGTGGGCACGGGAATCGTCAAGAGCGTCGTCAACGTATTCGGCGCGAAGCAG | 960
ptnpf1_2.8_allele1 | TTGTGGCCCGTGGGCACGGGAATCGTCAAGAGCGTCGTCAACGTATTCGGCGCGAAGCAG | 960
ptnpf1_2.8_allele2 | TTGTGGCCCGTGGGCACGGGAATCGTCAAGAGCGTCGTCAACGTATTCGGCGCGAAGCAG | 960

PtNPF1_wt | TTTCATCCACTTCTCCAGGCCTCCATGATTGAATCCTACTACGTCAAATTCTACATGTGC | 1020
ptnpf1_2.8_allele1 | TTTCATCCACTTCTCCAGGCCTCCATGATTGAATCCTACTACGTCAAATTCTACATGTGC | 1020
ptnpf1_2.8_allele2 | TTTCATCCACTTCTCCAGGCCTCCATGATTGAATCCTACTACGTCAAATTCTACATGTGC | 1020

PtNPF1_wt | ATCAATATTGGAGCCTTGGTCGGGGGTGTACTGGTACCCTTGCTGGCACAACACAACGTA | 1080
ptnpf1_2.8_allele1 | ATCAATATTGGAGCCTTGGTCGGGGGTGTACTGGTACCCTTGCTGGCACAACACAACGTA | 1080
ptnpf1_2.8_allele2 | ATCAATATTGGAGCCTTGGTCGGGGGTGTACTGGTACCCTTGCTGGCACAACACAACGTA | 1080

PtNPF1_wt | ACCCTCGCCTATTTCGTACCAGTCGGCATGCTGTCCTTGG------------------- | 1120
ptnpf1_2.8_allele1 | ACCCTCGCCTATTTCGTACCAGTCGGCATGCTGTCCTTGG------------------- | 1120
ptnpf1_2.8_allele2 | ACCCTCGCCTATTTCGTACCAGTCGGCATGCTGTCCTTGG------------------- | 1120

PtNPF1_wt | TCTGCTTTTATCTGCCCGGACTCACTCTCATTGCCCTCACAACCATACCGGGCGTTTTGG | 1260
ptnpf1_2.8_allele1 | TCTGCTTTTATCTGCCCGGACTCACTCTCATTGCCCTCACAACCATACCGGGCGTTTTGG | 1260
ptnpf1_2.8_allele2 | TCTGCTTTTATCTGCCCGGACTCACTCTCATTGCCCTCACAACCATACCGGGCGTTTTGG | 1260

PtNPF1_wt | GAGATACTTTTAACCGTCGCGCCCTTTCAATTGGTCTCCTCGTCCTTGTGGKACMGGGSCC | 1320
ptnpf1_2.8_allele1 | GAGATACTTTTAACCGTCGCGCCCTTTCAATTGGTCTCCTCGTCCTTGTGGKACMGGGSCC | 1320
ptnpf1_2.8_allele2 | GAGATACTTTTAACCGTCGCGCCCTTTCAATTGGTCTCCTCGTCCTTGTGGKACMGGGSCC | 1320

PtNPF1_wt | ACCTAAAGACCCCAGTGCCACACTGGCGGCCAATATCTGCCGTCGTCGACCCCCCAAGCG | 1440
ptnpf1_2.8_allele1 | ACCTAAAGACCCCAGTGCCACACTGGCGGCCAATATCTGCCGTCGTCGACCCCCCAAGCG | 1440
ptnpf1_2.8_allele2 | ACCTAAAGACCCCAGTGCCACACTGGCGGCCAATATCTGCCGTCGTCGACCCCCCAAGCG | 1440

PtNPF1_wt | TCCCGCCTACACCCCTCCCAACACGGATTCCATCGGACTGGACATGATCTTTCGCATCAG | 1500
ptnpf1_2.8_allele1 | TCCCGCCTACACCCCTCCCAACACGGATTCCATCGGACTGGACATGATCTTTCGCATCAG | 1500
ptnpf1_2.8_allele2 | TCCCGCCTACACCCCTCCCAACACGGATTCCATCGGACTGGACATGATCTTTCGCATCAG | 1500

PtNPF1_wt | CTTACTCATCATACCCTTCAACATTGCCTACTCCCAAATGGCCACCACCTTTATTGTCCA | 1560
ptnpf1_2.8_allele1 | CTTACTCATCATACCCTTCAACATTGCCTACTCCCAAATGGCCACCACCTTTATTGTCCA | 1560
ptnpf1_2.8_allele2 | CTTACTCATCATACCCTTCAACATTGCCTACTCCCAAATGGCCACCACCTTTATTGTCCA | 1560

PtNPF1_wt | AGGCACTGTTATGGAAAAAGCCTTTGGCTGGATCGACGCCGCCTGTATGAACAATGCCGA | 1620
ptnpf1_2.8_allele1 | AGGCACTGTTATGGAAAAAGCCTTTGGCTGGATCGACGCCGCCTGTATGAACAATGCCGA | 1620
ptnpf1_2.8_allele2 | AGGCACTGTTATGGAAAAAGCCTTTGGCTGGATCGACGCCGCCTGTATGAACAATGCCGA | 1620

PtNPF1_wt | CGCCGTCAGTGTCCTGCTCTTTGGCTACCTGATCGGTTCGCAGTTCTATCCCGCTCTCGC | 1680
ptnpf1_2.8_allele1 | CGCCGTCAGTGTCCTGCTCTTTGGCTACCTGATCGGTTCGCAGTTCTATCCCGCTCTCGC | 1680
ptnpf1_2.8_allele2 | CGCCGTCAGTGTCCTGCTCTTTGGCTACCTGATCGGTTCGCAGTTCTATCCCGCTCTCGC | 1680

PtNPF1_wt | GAATCGCGGTATCCGAATACCCACCACGTACAAGTTTGCGATTGGTTCCGGTCTCGGGGC | 1740
ptnpf1_2.8_allele1 | GAATCGCGGTATCCGAATACCCACCACGTACAAGTTTGCGATTGGTTCCGGTCTCGGGGC | 1740
ptnpf1_2.8_allele2 | GAATCGCGGTATCCGAATACCCACCACGTACAAGTTTGCGATTGGTTCCGGTCTCGGGGC | 1740

PtNPF1_wt | ACTGGCCATTGCGTGGGCCCTATTGGTAGAAGTCCTGATTCACGACCGGTTCGAGTCCAC | 1800
ptnpf1_2.8_allele1 | ACTGGCCATTGCGTGGGCCCTATTGGTAGAAGTCCTGATTCACGACCGGTTCGAGTCCAC | 1800
ptnpf1_2.8_allele2 | ACTGGCCATTGCGTGGGCCCTATTGGTAGAAGTCCTGATTCACGACCGGTTCGAGTCCAC | 1800

PtNPF1_wt | CGGAAAGCGCGTCAGCGTCCTCTGGCAAACCGTATCGTACGTGTTAATTGGTGCGGGCGA | 1860
ptnpf1_2.8_allele1 | CGGAAAGCGCGTCAGCGTCCTCTGGCAAACCGTATCGTACGTGTTAATTGGTGCGGGCGA | 1860
ptnpf1_2.8_allele2 | CGGAAAGCGCGTCAGCGTCCTCTGGCAAACCGTATCGTACGTGTTAATTGGTGCGGGCGA | 1860

PtNPF1_wt | GATCTTTGCCGCTCCGGCCCTACATCGGACTTGGCTCCTGATTTACCGGCGCCCGAAGAA | 1920
ptnpf1_2.8_allele1 | GATCTTTGCCGCTCCGGCCCTACATCGGACTTGGCTCCTGATTTACCGGCGCCCGAAGAA | 1920
ptnpf1_2.8_allele2 | GATCTTTGCCGCTCCGGCCCTACATCGGACTTGGCTCCTGATTTACCGGCGCCCGAAGAA | 1920
PtNPF1_wt               ACCGCTACTCCCTCGCCGCTGCTTGCCATTGGCACTCCACCGACCCGGACGCGACATTCC 300
ptnpf1_2.9_allele1      ACCGCTACTCCCTCGCCGCTGCTTGCCATTGGCACTCCACCGACCCGGACGCGACATTCC 300
ptnpf1_2.9_allele2

PtNPF1_wt               CCTTCGGCGTCTCCCGCACCGCAAACCAAGCGTGCGGCTGCCGTCGTCCCGAGTGTGCCC 360
ptnpf1_2.9_allele1      CCTTCGGCGTCTCCCGCACCGCAAACCAAGCGTGCGGCTGCCGTCGTCCCGAGTGTGCCC 360
ptnpf1_2.9_allele2

PtNPF1_wt               CCCTCGCGGGCTTCCTCACGCGCACCGTCACCCATTGGCTCGGACGCTTCCGACGAATTG 420
ptnpf1_2.9_allele1      CCCTCGCGGGCTTCCTCACGCGCACCGTCACCCATTGGCTCGGACGCTTCCGACGAATTG 420
ptnpf1_2.9_allele2

PtNPF1_wt               CCCAACAACGTGTGTACCGAAGAACAACTCTTGTTGCTGCCCGAAGAATGTTTGAGTGGC 480
ptnpf1_2.9_allele1      CCCAACAACGTGTGTACCGAAGAACAACTCTTGTTGCTGCCCGAAGAATGTTTGAGTGGC 480
ptnpf1_2.9_allele2

PtNPF1_wt               CACCATGTGCGTCCGCTACGCCACGTCGATGAAGAAGGCAACGAAAGTTTCTACCACCTG 540
ptnpf1_2.9_allele1      CACCATGTGCGTCCGCTACGCCACGTCGATGAAGAAGGCAACGAAAGTTTCTACCACCTG 540
ptnpf1_2.9_allele2

PtNPF1_wt               CGTCCCATGTTCTATTCCGTCATTTTTATCCTGCTTGTGGAACTGCTCGAGCGATTTTCC 600
ptnpf1_2.9_allele1      CGTCCCATGTTCTATTCCGTCATTTTTATCCTGCTTGTGGAACTGCTCGAGCGATTTTCC 600
ptnpf1_2.9_allele2

PtNPF1_wt               TTTTACGGCATTAATTACACACAAACGTCCTACCTGACCGGCTCCTACAATCGCGATTGG 660
ptnpf1_2.9_allele1      TTTTACGGCATTAATTACACACAAACGTCCTACCTGACCGGCTCCTACAATCGCGATTGG 660
ptnpf1_2.9_allele2

PtNPF1_wt               AACGCCGACATGGAAGCCGTCGACGCTTCCACCTACGTTTCCATTTCGGTTGCCGTTGCC 720
ptnpf1_2.9_allele1      AACGCCGACATGGAAGCCGTCGACGCTTCCACCTACGTTTCCATTTCGGTTGCCGTTGCC 720
ptnpf1_2.9_allele2

PtNPF1_wt               TACACCTCTCCCTTTCTCGGGGCCTATCTAGCCGACGCCATTCTCGGCGACTACGGATCC 780
ptnpf1_2.9_allele1      TACACCTCTCCCTTTCTCGGGGCCTATCTAGCCGACGCCATTCTCGGCGACTACGGATCC 780

PtNPF1_wt               CTCTTTGTCGGATCTCTCTGCTTTTATCTGCCCGGACTCACTCTCATTGCCCTCACAACC 840
ptnpf1_2.9_allele1      CTCTTTGTCGGATCTCTCTGCTTTTATCTGCCCGGACTCACTCTCATTGCCCTCACAACC 840
ptnpf1_2.9_allele2

PtNPF1_wt               ATACCGGGCGTTTTGGGAGATACTTTTAACCGTCGCGCCCTTTCAATTGGTCTCCTCGTC 900
ptnpf1_2.9_allele1      ATACCGGGCGTTTTGGGAGATACTTTTAACCGTCGCGCCCTTTCAATTGGTCTCCTCGTC 900
ptnpf1_2.9_allele2

PtNPF1_wt               TTGTGGCCCGTGGGCACGGGAATCGTCAAGAGCGTCGTCAACGTATTCGGCGCGAAGCAG 960
ptnpf1_2.9_allele1      TTGTGGCCCGTGGGCACGGGAATCGTCAAGAGCGTCGTCAACGTATTCGGCGCGAAGCAG 960
ptnpf1_2.9_allele2

PtNPF1_wt               TTTCATCCACTTCTCCAGGCCTCCATGATTGAATCCTACTACGTCAAATTCTACATGTGC 1020
ptnpf1_2.9_allele1      TTTCATCCACTTCTCCAGGCCTCCATGATTGAATCCTACTACGTCAAATTCTACATGTGC 1020
ptnpf1_2.9_allele2

PtNPF1_wt               ATCAATATTGGAGCCTTGGTCGGGGGTGTACTGGTACCCTTGCTGGCACAACACAACGTA 1080
ptnpf1_2.9_allele1      ATCAATATTGGAGCCTTGGTCGGGGGTGTACTGGTACCCTTGCTGGCACAACACAACGTA 1080
ptnpf1_2.9_allele2

PtNPF1_wt               ACCCTCGCCTATTTCGTACCAGTCGGCATGCTGTCCTTGGGAGTCGGTCTCTTTGCCGTG 1140
ptnpf1_2.9_allele1      ACCCTCGCCTATTTCGTACCAGTCGGCATGCTGTCCTTGGGAGTCGGTCTCTTTGCCGTG 1140
ptnpf1_2.9_allele2

PtNPF1_wt               GGATCCAAACGCTACGTCTGCGACCAACCTAAAGACCCCAGTGCCACACTGGCGGCCAAT 1200
ptnpf1_2.9_allele1      GGATCCAAACGCTACGTCTGCGACCAACCTAAAGACCCCAGTGCCACACTGGCGGCCAAT 1200
ptnpf1_2.9_allele2

PtNPF1_wt               ATCTGCCGTCGTCGACCCCCCAAGCGTCCCGCCTACACCCCTCCCAACACGGATTCCATC 1260
ptnpf1_2.9_allele1      ATCTGCCGTCGTCGACCCCCCAAGCGTCCCGCCTACACCCCTCCCAACACGGATTCCATC 1260
ptnpf1_2.9_allele2

PtNPF1_wt               GGACTGGACATGATCTTTCGCATCAGCTTACTCATCATACCCTTCAACATTGCCTACTCC 1320
ptnpf1_2.9_allele1      GGACTGGACATGATCTTTCGCATCAGCTTACTCATCATACCCTTCAACATTGCCTACTCC 1320
ptnpf1_2.9_allele2
ptnpf2_1.16_allele2   TTTGGAGATTTACTGAGTGGAATCTTGTATTCCACCGTGTTTGCGAATATGAATCGAGCG  619
************************************************************
PtNPF2_wt                AAAATCATGCATACCTGTGCCTTGCTTATGCTGTGTAACTTGGGATTATTTGCGCTCGTG 1860
ptnpf2_1.16_allele1      AAAATCATGCATACCTGTGCCTTGCTTATGCTGTGTAACTTGGGATTATTTGCGCTCGTG 1855
ptnpf2_1.16_allele2      AAAATCATGCATACCTGTGCCTTGCTTATGCTGTGTAACTTGGGATTATTTGCGCTCGTG  679
************************************************************
PtNPF2_wt                GTTCGGTGGTGGGAACGTCGCGAAGTGCACGATTTAAGGCGTTTACAGTCCCTCCAGGGG 1920
ptnpf2_1.16_allele1      GTTCGGTGGTGGGAACGTCGCGAAGTGCACGATTTAAGGCGTTTACAGTCCCTCCAGGGG 1915
ptnpf2_1.16_allele2      GTTCGGTGGTGGGAACGTCGCGAAGTGCACGATTTAAGGCGTTTACAGTCCCTCCAGGGG  739
******************************************************************************
PtNPF2_wt                CTGGAACTACGAGAAGAGCGAAGAATGATTTGA 1953
ptnpf2_1.16_allele1      CTGGAACTACGAGAAGAGCGAAGAATGATTTGA 1948
ptnpf2_1.16_allele2      CTGGAACTACGAGAAGAGCGAAGAATGATTTGA        772
## Appendix C. Supplementary materials relative to Chapter 4

### Supplementary Table S4.1

List of vital organelle trackers available for *P. tricorntum*, including references of studies where staining was optimised for diatoms.

<table>
<thead>
<tr>
<th>Organelle</th>
<th>Tracker name</th>
<th>Excitation wavelength (nm)</th>
<th>Emission wavelength (nm)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleus</td>
<td>Bisbenzimide H 33342 (Hoechst 33342; Sigma, MO, USA)</td>
<td>405</td>
<td>425-475</td>
<td>(Costa et al., 2013)</td>
</tr>
<tr>
<td>Mitochondrion</td>
<td>MitoTracker Orange CMTMRos (Molecular probes, Eugene, OR, USA)</td>
<td>561</td>
<td>570-620</td>
<td>(Tanaka et al., 2015)</td>
</tr>
<tr>
<td>Plasmamembrane - cytoskeleton</td>
<td>FITC-phalloidin (Molecular Probes, Eugene, USA)</td>
<td>488</td>
<td>525</td>
<td>(Galas et al., 2021)</td>
</tr>
<tr>
<td>Golgi</td>
<td>Golgi-Tracker Red (Beyotime, Beijing, China)</td>
<td>589</td>
<td>610-650</td>
<td>(Shao et al., 2019)</td>
</tr>
<tr>
<td>ER</td>
<td>ER-Tracker Blue-White DPX (Invitrogen)</td>
<td>375</td>
<td>550-640</td>
<td>(Shao et al., 2019)</td>
</tr>
<tr>
<td>Vacuole</td>
<td>Yeast Vacuole Membrane Marker MDY-64 (ThermoFisher Scientific)</td>
<td>418</td>
<td>490-520</td>
<td>(Huang et al., 2016)</td>
</tr>
<tr>
<td>Lipid bodies</td>
<td>Nile Red N1142 (Invitrogen)</td>
<td>515</td>
<td>585</td>
<td>(Wong and Franz, 2013)</td>
</tr>
<tr>
<td></td>
<td>BODIPY 505/515 (4,4-Difluoro-1,3,5,7-Tetramethyl-4-Bora-3a,4a-Diaza-s-Indacene) (Invitrogen)</td>
<td>505</td>
<td>515</td>
<td>(Wong and Franz, 2013)</td>
</tr>
</tbody>
</table>
Supplementary Figure S4.1

Correlation of cell concentration, measured through Malassez chambers, and in vivo chlorophyll a fluorescence, determined by using multifunctional monochromator-based microplate reader (Infinite™ M1000 Pro; Tecan), with excitation and emission wavelength set at 662 and 685 nm, respectively.

Supplementary Table S4.2

Number of experiments performed, and number of cells observed through confocal microscopy for morphological parameters per each condition and per each strain. 1KO indicates the *Ptnpf1* knock-out mutant 1KO 2.9, while 1OE indicated the PtNPF1-YFP overexpressing strain 1OE 4.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Number of experiments</th>
<th>Strain</th>
<th>Condition</th>
<th>Total number of cells analysed</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT: N starvation for 4 days</td>
<td>3</td>
<td>WT</td>
<td>Normal condition</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N starvation for 4 days</td>
<td>29</td>
</tr>
<tr>
<td>WT, 1KO and 1OE:</td>
<td>3</td>
<td>WT</td>
<td>Normal condition</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N starvation for 2 days</td>
<td>27</td>
</tr>
<tr>
<td>N starvation for 2, 4 and 7 days</td>
<td>1KO 2.9</td>
<td>Normal condition</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td>---------</td>
<td>-----------------</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N starvation for 2 days</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N starvation for 4 days</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N starvation for 7 days</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1OE 4</td>
<td>Normal condition</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N starvation for 2 days</td>
<td>29</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N starvation for 4 days</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N starvation for 7 days</td>
<td>22</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix D. Supplementary Materials relative to Chapter 5

Supplementary Figure S5.1

Photosynthetic parameters calculated during the low pH shift experiment on *P. tricornutum* wild-type, two PtNPF2-YFP overexpressing strains and two *Ptnpf2* mutants. A) Relative electron transport (rETR - μmol photons m$^{-2}$ s$^{-1}$) vs. light intensity (E - μmol photons m$^{-2}$ s$^{-1}$); B) Non-Photochemical Quenching (NPQ) vs. light intensity; C) NPQ relaxation vs. time in darkness.
Maximum fluorescence, used as a proxy for growth, of *P. tricornutum* wild-type, two PtNPF2-YFP overexpressing strains and two *Ptnpf2* mutants after the shift to A) normal pH and B) low pH. Photosystem II size was calculated, respectively for shift to C) normal
pH and D) low pH. Relative reaction centers content in the Photosystem II was also calculated at different time points, for the shift to E) normal pH and F) low pH. Error bars represent standard deviations of three biological replicates. * indicates $p < 0.05$; ** for $p < 0.01$ and *** for $p < 0.001$.

Supplementary Figure S5.3

$\alpha$ and $E_k$ values for *P. tricornutum* wild-type, two PtNPF2-YFP overexpressing strains and two *ptnpf2* mutants after the shift to normal pH and low pH. $\alpha$ values after 3, 24 and 48 hours from the shift to normal pH (A) and low pH (B). $E_k$ values after 3, 24 and 48 hours from the shift to normal pH (C) and low pH (D). Error bars represent standard deviations of three biological replicates. * indicates $p < 0.05$ and ** for $p < 0.01$. 

322
Supplementary Figure S5.4

E50\textsubscript{NPQ} and nNPQ values for \textit{P. tricornutum} wild-type, two PtNPF2-YFP overexpressing strains and two \textit{Ptnpf2} mutants after the shift to normal pH and low pH. E50\textsubscript{NPQ} values after 3, 24 and 48 hours from the shift to normal pH (A) and low pH (B). nNPQ values after 3, 24 and 48 hours from the shift to normal pH (C) and low pH (D). Error bars represent standard deviations of three biological replicates. * indicates $p < 0.05$; ** for $p < 0.01$ and *** for $p < 0.001$. 
Supplementary Figure S5.5

NPQ relaxation components for *P. tricornutum* wild-type, two PtNPF2-YFP overexpressing strains and two *Ptnpf2* mutants after the shift to normal pH. NPQ$_{\text{max}}$ values after 3, 24 and 48 hours from the shift to normal pH (A) and low pH (B). D, E, F) qE$_{\text{max}}$ values after 3, 24 and 48 hours from the shift to normal pH (C) and low pH (D). qI values after 3, 24 and 48 hours from the shift to normal pH (E) and low pH (F). Error bars represent standard deviations of three biological replicates. * indicates $p < 0.05$; ** for $p < 0.01$ and *** for $p < 0.001$. 
Supplementary Figure S5.6

$F_v/F_m$ recovery for *P. tricornutum* wild-type, two PtNPF2-YFP overexpressing strains and two *Ptnpf2* mutants after the shift to normal and low pH. A) $F_v/F_m$ recovery values after 3, 24 and 48 hours from the shift to normal pH. B) $F_v/F_m$ recovery values after 3, 24 and 48 hours from the shift to low pH. Error bars represent standard deviations of three biological replicates. * indicates $p < 0.05$ and ** for $p < 0.01$. 
A  
**F<sub>v</sub>/F<sub>m</sub> recovery - pH 8**

- WT
- 2KO 1.15
- 2KO 1.16
- 2OE 2
- 2OE 3

B  
**F<sub>v</sub>/F<sub>m</sub> recovery - pH 7**

- WT
- 2KO 1.15
- 2KO 1.16
- 2OE 2
- 2OE 3