UNIVERSITY OF SOUTHAMPTON

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

SCHOOL OF OCEAN AND EARTH SCIENCES

NUTRIENT UTILISATION BY *TRICHODESMIUM*

CHARACTERISATION OF MOLECULAR AND PHYSIOLOGICAL PROCESSES

by

DESPO POLYVIOU

Thesis for the degree of Doctor of Philosophy
August 2016
Despo Polyviou:

Nutrient utilisation by *Trichodesmium*
Characterisation of Molecular and Physiological Processes

**Supervised by:**
Prof. Thomas S. Bibby, Prof. C. Mark Moore,
Dr Andrew Hitchcock, Dr Martha Gledhill and Prof. Matt Mowlem
I dedicate this book to the three people who have made me who I am today:

My brother, who taught me that shoes need to be taken for walks and that you can be anyone you like to be (be that a fairy tale character or a cool animal.

My mum, the strongest person I know, who taught me what love is, what patience is, and who believed in me when no one else did (not even me).

My dad, who inspired my love for the sea, and taught me to never cease searching both inside and around me.
“I ca’n’t believe that!” said Alice.

“Ca’n’t you?” the Queen said in a pitying tone.

“Try again: draw a long breath, and shut your eyes.”

Alice laughed. “There’s no use trying,” she said:

“one ca’n’t believe impossible things.”

“I daresay you haven’t had much practice,” said the Queen. “When I was your age, I always did it for half-an-hour a day.

Why, sometimes I’ve believed as many as six impossible things before breakfast.”

Through the Looking-Glass
Lewis Carroll
The activity of photosynthetic cyanobacteria capable of nitrogen (N\textsubscript{2}) fixation (diazotrophs) strongly influences oceanic primary production and global biogeochemical cycles. The niche of these organisms extends mainly across low latitude oligotrophic oceans, largely deficient in nitrate, where they introduce 'new' nitrogen (N) to the system. In these regions the abundant marine cyanobacterium \textit{Trichodesmium} spp. accounts for a significant proportion of the fixed N flux.

Despite fixation of N, the availability of phosphorus (P) and iron (Fe) remain a constraint to the activity and biogeography of diazotrophs. The genome of \textit{Trichodesmium} has therefore been shaped to provide intricate adaptive strategies optimising growth under both P and Fe depletion. Characterisation of these strategies can provide information that will enhance the understanding of the organism's biogeography in the contemporary and future ocean.

In this work, molecular and physiological techniques are employed to study nutrient uptake pathways, and the metabolic response of \textit{Trichodesmium erythraeum} IMS101 (\textit{Trichodesmium} hereafter) to nutrient limitation. The current lack of an established system for genetic manipulation of this organism inhibits direct functional characterisation of proteins. To circumvent this, the model cyanobacteria \textit{Synechocystis} sp. PCC 6803 (\textit{Synechocystis} hereafter) is used as a vehicle for the heterologous expression of \textit{Trichodesmium} genes.

Using this technique, the suggested contribution of \textit{Trichodesmium} to an emerging oceanic P redox cycle is first explored. A four-gene cluster (ptxABCD), that encodes a putative ABC transporter (ptxABC) and NAD-dependent dehydrogenase (ptxD), is demonstrated to be responsible for the organism's ability to utilise the reduced inorganic compound phosphite. The presence and expression of this gene cluster is also confirmed in diverse field metagenomic and metatranscriptomic datasets further confirming its role in \textit{Trichodesmium} species.

Pathways of Fe utilisation are also investigated. Through heterologous expression the function of a currently employed Fe stress biomarker, protein Tery\_3377 (IdiA), which is homologous to both Fe\textsuperscript{3+} transporters (FutA2-like) and intracellular proteins with protective function under Fe stress (FutA1-like), is elucidated. Fusing the signal sequence of this protein to GFP revealed its periplasmic localisation, and its expression in \textit{Synechocystis} mutants of both futA1 and futA2 paralogues further supported involvement in Fe\textsuperscript{3+} uptake, providing evidence for its function as an Fe transporter in \textit{Trichodesmium}.

Finally, a physiological experiment was performed to determine the significance of direct physical contact with Saharan desert dust for acquisition of Fe by \textit{Trichodesmium}. It is demonstrated that cell surface processes are fundamental in dust-Fe utilisation by this organism and transcriptomic analysis identifies a number of unique genes regulated under different Fe and dust regimes including putative cell-surface proteins not previously studied in \textit{Trichodesmium}.

Combined, these studies have revealed a diverse array of molecular and physiological strategies potentially employed by \textit{Trichodesmium} to survive and thrive on the ephemeral supplies of nutrients encountered in oligotrophic oceans, an attribute that facilitates its significant contribution to biogeochemical cycles.
# Table of Contents

Chapter 1 .................................................................................................................. 1

1.1 The importance of marine primary production .................................................. 1

1.2 Nutrients in the ocean ......................................................................................... 3

1.2.1 Phosphorus ..................................................................................................... 6

1.2.2 Iron .................................................................................................................. 8

1.2.3 Paradigms of nutrient limitation ................................................................. 9

1.3 Marine Diazotrophy ......................................................................................... 10

1.3.1 The players .................................................................................................... 10

1.3.2 Oxygen avoidance ....................................................................................... 16

1.4 Living with less - lessons from marine microorganisms .................................. 18

1.4.1 Cellular strategies to cope with P stress ...................................................... 18

1.4.2 Cellular strategies to cope with Fe stress ................................................. 22

1.4.3 Trophic life strategies - Genomic minimalism or maximalism? ............ 27

1.5 Omics and Detection of stress responses ....................................................... 28

1.6 Significance of molecular data in ocean science ............................................. 31

1.7 Thesis plan ......................................................................................................... 32

Chapter 2 .................................................................................................................. 35

2.1 Strains and growth conditions ........................................................................ 35

2.1.1 *E. coli* ......................................................................................................... 35

2.1.2 *Trichodesmium* ......................................................................................... 35

2.1.3 *Synechocystis* .......................................................................................... 39

2.2 Physiological measurements ........................................................................... 40

2.2.1 Photosynthetic physiology ........................................................................ 40

2.2.2 Growth ......................................................................................................... 41

2.2.3 Chlorophyll *a* ............................................................................................ 42

2.2.4 Statistical analysis ....................................................................................... 42

2.3 Methods for molecular analysis ...................................................................... 42

2.3.1 Nucleic acid extractions and quality control ............................................ 42

2.3.2 DNase Treatment and cDNA synthesis .................................................... 44
List of Figures

Figure 1.1: Phytoplankton activity and downstream interactions with the ocean system. .... 2
Figure 1.2: The link between nutrients and primary production in the ocean. ....................... 5
Figure 1.3: Interactions of phosphorus and iron forms with the biotic and abiotic environment in the ocean. ........................................................................................................... 7
Figure 1.4: Marine diazotrophy in nutrient limited oceans. ........................................................ 11
Figure 1.5: Trichodesmium morphology. ..................................................................................... 14
Figure 1.6: Strategies of N2 fixation-photosynthesis segregation .............................................. 17
Figure 1.7: The Phosphorus and Iron uptake and utilisation pathways in microorganism ...21
Figure 1.8: The cellular information flow and interactions with the environment. ................. 29
Figure 2.1: Generation of deletion mutants. .............................................................................. 48
Figure 2.2: pFLAG vector map. ................................................................................................ 50
Figure 2.3: Generation of the GFP tagged signal sequence strains ........................................... 52
Figure 3.1: Growth and physiology of Trichodesmium in different P regimes ....................... 67
Figure 3.2: The arrangement of the ptxABCD genes in Trichodesmium ................................. 68
Figure 3.3: Gene expression analysis in response to P source ................................................... 70
Figure 3.4: Expression of genes along the ptx and phn clusters .............................................. 72
Figure 3.5: Confirmation of strain generation ............................................................................ 74
Figure 3.6: Growth of Synechocystis with phosphate requires all four ptx genes ................. 75
Figure 3.7: Strains of Synechocystis grow unevenly on phosphate ........................................... 76
Figure 4.1: Current model of the localisation of Synechocystis Fut paralogues ....................... 83
Figure 4.2: Maximum likelihood phylogenetic analysis of Tery_3377 homologues from selected bacteria ......................................................................................................................... 86
Figure 4.3: The GFP tagged-signal sequence strains ............................................................... 90
Figure 4.4: Localization of signal sequence-GFP constructs in Synechocystis ....................... 91
Figure 4.5: Identification of GFP expression in Synechocystis ................................................ 93
Figure 4.6: Localization of the Trichodesmium Tery_3377 signal sequence-GFP constructs in Synechocystis ............................................................................................................. 93
Figure 4.7: Synechocystis ΔfutA1 and ΔfutA2 deletion mutants were generated ....................... 94
Figure 4.8: Physiology of FutA mutant strains in Synechocystis ............................................. 95
Figure 4.9: Expression of FutA1 and FutA2 in Synechocystis deletion and complemented mutants ......................................................................................................................... 96
Figure 4.10: Photosynthetic physiology of strains under Fe stress. ......................................... 96
Figure 4.11: Growth rate of strains under Fe stress. ............................................................... 99
Figure 4.12: Size Exclusion Chromatography (SEC) of Tery_3377 ............................................. 101
Figure 4.13: Purification and crystallisation of Tery_3377 ......................................................... 102
Figure 5.1: The experimental design ....................................................................................... 110
Figure 5.2: Experimental design to assess possible differential dissolution of nutrients from dust enclosed or not enclosed in dialysis tubing ......................................................... 111
Figure 5.3: Assessment of potential dialysis tubing effects when included in the experiment .......................................................................................................................... 112
Figure 5.4: Assessment of potential inhibitory effect of the dialysis tubing on abiotic dust-chemical dissolution. ................................................................. 113
Figure 5.5: Growth of *Trichodesmium* under different Fe regimes. ........................... 114
Figure 5.6: *Trichodesmium* filament physiology under different Fe regimes ......... 116
Figure 5.7: *Trichodesmium* photosynthetic efficiency (*F_{m}/F_{n*}) under different Fe regime. 117
Figure 5.8: Trichodesmium functional absorption cross section of PSII (*σ_{PSII}*) during growth under different Fe regimes ................................................................. 118
Figure 5.9: Gene numbers and GO annotation coverage .................................................. 120
Figure 5.10: Distribution of the number of genes expressed across different levels of reads/gene ......................................................................................... 120
Figure 5.11: Gene expression clustering across treatments and replicates ....................... 121
Figure 5.12: Comparison of differential expression across the experimental treatments . 122
Figure 5.13: Clustering demonstrates patterns of differential expression across all genes and treatments. ..................................................................................... 123
Figure 5.14: Validation of the RNAseq results using quantitative RT-PCR (qPCR). ........ 126
Figure 5.15: GO enrichment in differentially expressed genes. ...................................... 128
Figure 5.16: The balance of positive and negative changes in gene expression ............ 131
Figure 5.17: Comparisons of transcriptomic changes due to additions of Dust and FeCl$_3$. 132
Figure 5.18: Iron uptake across the inner and outer cell membranes. .............................. 136
Figure 5.19: Gene expression of various components of putative Fe uptake pathways in *Trichodesmium* ................................................................. 140
Figure 5.20: Changes in transcription of genes involved in the process of photosynthesis. .............................................................................................. 142
Figure 5.21: Transcriptomic response of subunits involved in the photosynthetic electron transport chain ................................................................. 146
Figure 5.22: Gene expression of putative outer membrane proteins ............................... 151
Figure 5.23: Representative expression across the *Tery_08xx* cluster .......................... 153
Figure 5.24: Representative expression profile along the *Trichodesmium* twintron arrangement. ................................................................. 156
Figure 5.25: Expression of the *Trichodesmium* twintron across the experimental treatments ................................................................. 157
Figure 5.26: Expression of the *Trichodesmium* diversity generating retroelement .... 158
Figure 6.1: Iron and Phosphorus uptake by *Trichodesmium*. ......................................... 169
Figure 6.2: The prevalence /expression of the ptx gene cluster in marine cyanobacteria *Trichodesmium* and *Prochlorococcus* ........................................... 171
Figure 6.3: The hypothesised response of the photosynthetic electron transport (PET) chain to Fe limitation in *Trichodesmium*. ........................................... 176
Figure A.1 Clustered expression patterns across the four experimental treatments .... 191
List of Tables

Table 1A: Characteristics of the major diazotrophic cyanobacteria groups ........................................... 12
Table 1B: Sources of iron accessible to *Trichodesmium* ................................................................. 25
Table 2A: Bacterial strains used for experiments .......................................................................................... 36
Table 2B: Modified (m) YBC-II media recipe ............................................................................................. 37
Table 2C: Modified YBC-II trace metal composition compared to standard YBC-II ............................. 38
Table 2D: BG-11 media recipe .................................................................................................................. 38
Table 2E: Fast Repetition Rate fluorometer settings ................................................................................... 40
Table 2F: Description of physiological parameters recorded ...................................................................... 41
Table 2G: RNA extracted- quality control and concentrations ................................................................. 44
Table 2H: Genes analysed by qPCR, amplicon and primer information ......................................................... 48
Table 2I: Plasmid generated in this study .................................................................................................... 50
Table 2J: Buffers used for Tery_3377 protein extraction ............................................................................ 57
Table 2K: Tery_3377 homologues included in the phylogenetic analysis .................................................... 61
Table 3A: The different forms and oxidation states of phosphorus in the ocean .................................... 64
Table 3B: Details of the *ptx* homologous genes in *Trichodesmium* ......................................................... 69
Table 3C: The expression of phosphite (*ptx*) and phosphonate (*phn*)-metabolism genes in *Trichodesmium* transcriptomes .................................................................................................................. 70
Table 3D: Putative Pho box promoter binding sequences associated with P-metabolism involved genes ......................................................................................................................................................... 73
Table 3E: Environmental prevalence of *ptx* genes in *Trichodesmium* .................................................... 78
Table 4A: Phenotypes previously reported in Synechocystis Δ*futA1* and Δ*futA2* mutants .................. 82
Table 4B: IdiA/FutA/Fbp homologues of selected bacteria ........................................................................ 85
Table 5A: Clustering of differentially expressed genes in 10 distinct patterns across experimental treatments ........................................................................................................................................................................ 124
Table 5B: The top20 differentially expressed genes presented is descending order of statistical significance (P-value) ........................................................................................................................................ 129
Table 5C: Expression of suggested biomarkers of iron limitation ............................................................. 134
Table 5D: The *Trichodesmium Tery_08xx* operon components ................................................................. 154
Table A1: Primer sequences used for genetic modification ..................................................................... 184
Table A2: The 50 top differentially expressed genes in the transcriptomic analysis ............................. 186
Table A3: The 50 most highly expressed genes in the transcriptomic analysis ........................................ 188
<table>
<thead>
<tr>
<th>Acronyms</th>
<th>Full Forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic local alignment search tool</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary deoxyribonucleic acid</td>
</tr>
<tr>
<td>Chl-α</td>
<td>Chlorophyll-α</td>
</tr>
<tr>
<td>DIP</td>
<td>Dissolved inorganic phosphorus</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNase</td>
<td>Deoxyribonuclease</td>
</tr>
<tr>
<td>DOP</td>
<td>Dissolved organic phosphorus</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothrietol</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>FRRf</td>
<td>Fast Repetition Rate fluorometer</td>
</tr>
<tr>
<td>gDNA</td>
<td>Genomic deoxyribonucleic acid</td>
</tr>
<tr>
<td>GLM</td>
<td>General linear model</td>
</tr>
<tr>
<td>HNLC</td>
<td>high nutrient low chlorophyll</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>kDa</td>
<td>Kilodalton</td>
</tr>
<tr>
<td>MIQE</td>
<td>Minimum information for publication of quantitaive real-time PCR experiments</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger ribonucleic acid</td>
</tr>
<tr>
<td>NADPH</td>
<td>Nicotinamide adenine dinucleotide phosphate</td>
</tr>
<tr>
<td>NCBI</td>
<td>National center for biotechnology information</td>
</tr>
<tr>
<td>NRQ</td>
<td>Normalised relative quantity</td>
</tr>
<tr>
<td>NTC</td>
<td>No template control</td>
</tr>
<tr>
<td>OP</td>
<td>Organophosphorus</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PIP</td>
<td>Particulate inorganic phosphorus</td>
</tr>
<tr>
<td>Pn</td>
<td>Phosphonate</td>
</tr>
<tr>
<td>POC</td>
<td>Particulate organic carbon</td>
</tr>
<tr>
<td>polyP</td>
<td>Polyphosphate</td>
</tr>
<tr>
<td>POP</td>
<td>Particulate organic phosphorus</td>
</tr>
</tbody>
</table>
PSI Photosystem I
PSII Photosystem II
qPCR Quantitative polymerase chain reaction
RFC Relative fold change
RNA Ribonucleic acid
rRNA Ribosomal ribonucleic acid
RT-PCR Reverse transcription polymerase chain reaction
ROS Reactive oxygen species
SEA Singular enrichment analysis
SOD Superoxide dismutase
I, Despo Polyviou declare that this thesis, and the work presented in it are my own

Nutrient utilisation by *Trichodesmium*—Characterisation of Molecular and Physiological Processes

I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

Parts of this work have been published as:

Signed: ______________

Date: ______________
Acknowledgements

I am determined that if someone was to rip off this page and place it on the opposite side of the scale to the rest of the book the scale would shift towards the side of the single page. Through it, I want to acknowledge the people that I travelled with along the path that lead to the point I stand now. For no matter how fulfilling it is to reach a destination, the road and the adventures you have with people along the way is what makes you, as a person and as a scientist.

First and foremost, I would like to express my gratitude to my supervisors: Thomas Bibby, my advisor and mentor for inspiring me to become a scientist and supporting me every step of the way. Having the special ability to distinguish when it was time for each, he guided me when I needed guidance and encouraged me to lead my own way when I was ready. Mark Moore, for his clarity of thought and good advice. Having him stand behind me always made me feel safe and confident in our research endeavours. Andrew Hitchcock, for teaching me everything I know from holding a pipette to making bacteria glow fluorescent green, and without whom this thesis would never have been possible.

Special thanks go to my examiners Maeve Lohan and James Murray for making my viva examination a pleasant experience and their useful corrections and comments on this work.

I also thank Martha Gledhill for although she had to migrate at the beginning of my PhD she first managed to pass on some of her infectious enthusiasm for chemistry, which sometimes makes you dance in the lab. I am also grateful to Ben Van Mooy for taking me out of the lab and teaching me some real oceanography.

Thanks to all that I had the luck to produce amazing science with. Joe Snow, the protein expert, whose thesis was incredibly useful while writing this book, and Alison Baylay, the DNA expert, who was patient with me and present whenever I needed her, either for scientific advice or to just have coffee and cake. I thank Moritz Machelett and Ivo Tews for teaching me how to make pretty pink crystals. I am also grateful for Mark Hopwood’s determination to convert me to chemistry. I enjoyed the stimulating discussions we shared in 166/09 and the sleepless nights of science in Crete.

I am indebted to Nicola Pratt who helped me get started and continued helping me thereafter and John Gittins for providing his expert advice when needed. I am also appreciative of all the people I had the pleasure to work next to in the lab including Andreas the single other member of the DnA team for simply being awesome, Nathan, Wladimir, Sophie, Tommy, Harry, Dok, and Annie.

I am grateful to James for taking me along on his adventures and showing me new things. He gave my life forward momentum and would be happy to know that towards the end of this book I started to take regular ‘fast walks’. Thank you also goes to my friends and especially Nina, Giang and Glaucia for the fun we had together and their support whenever it was needed, Nicole for always being there, Iordanis for bringing a piece of home to the office, and all my officemates for making NOCS a more pleasant place to work.

Finally, I thank my family and especially my grandparents for passing me their thirst for knowledge, and encouraging me to take advantage of all the opportunities that they themselves were never offered.

Work for this PhD was supported by the Graduate School of the National Oceanography Centre Southampton, the Natural Environmental Research Council and the A.G. Leventis Foundation.
Chapter 1
Introduction

1.1 The importance of marine primary production

Marine microbial life is of foremost importance in catalysing oceanic biogeochemical cycles which are bi-directionally linked to the global climate. Phytoplankton accounts for about half of the global annual primary production through photosynthetic fixation (Box 1) of ~48.5 Pg of carbon every year (Field et al. 1998; Falkowski et al. 2004). Via a process referred to as the marine biological carbon pump (Fig. 1.1) fixed organic carbon is exported from surface waters to depth, with a fraction subsequently stored within marine sediments. Since climate changes over glacial/interglacial periods correlate to the atmospheric CO$_2$ concentrations, which can be explained by the activity of the biological carbon pump (Sigman & Boyle 2000), it is vital to understand the efficiency of, and limiting factors controlling these process in the context of rapidly changing climatic conditions. As a more direct consequence, perturbations to oceanic primary productivity can also affect fisheries production (Blanchard et al. 2012) and consequently food security.
Figure 1.1: Phytoplankton activity and downstream interactions with the ocean system. The biological carbon pump is powered by growth of autotrophic organisms through atmospheric CO\(_2\) drawdown and carbon fixation. Nitrogen (N\(_2\)) fixation enhances primary production either directly through growth of photosynthetic diazotrophs or through the release of fixed N in the surrounding environment. Particulate organic carbon (POC) is consequently exported from the euphotic zone to the ocean sediment. Some POC escapes from this downward flux and is recycled, becoming available to support regenerated production. (Figure adapted from Herndl & Reinthaler 2013).

Although marine macrophytes account for a significant proportion of coastal photosynthetic activity, in the open ocean CO\(_2\) fixation is dominated by unicellular autotrophs (Duarte 2009, p.10). These are a diverse group composed primarily of the eukaryotic microalgae groups: coccolithophores, dinoflagellates and diatoms, as well as prokaryotic marine cyanobacteria (Simon et al. 2009). Ancestors of the latter were the first organisms capable of oxygenic photosynthesis, a process that lead to the oxygenation of the oceans at the end of the Pre-Cambrian period (Kaufman 2014). They include *Synechococcus* and *Prochlorococcus*, major contributors to oceanic primary production (Liu et al. 1997; Partensky et al. 1999; Scanlan 2003), and nitrogen fixing organisms such as *Trichodesmium* and *Crocosphaera* (Carpenter et al. 1992).
1.2 Nutrients in the ocean

Vast areas of the ocean, referred to as oligotrophic, are depleted in the macronutrients nitrogen (N) and phosphorus (P) and therefore limit primary production (Fig. 1.2) (Moore et al. 2013). A group of prokaryotic microorganisms, referred to as diazotrophs, have developed the ability to take up inorganic dinitrogen (N$_2$) from their environment and reduce it to two molecules of ammonia (NH$_3$) (Box 2) which can be consequently utilised for growth and to carry out essential metabolic processes. This drives CO$_2$ fixation by
autorophic diazotrophs directly, but also ‘new nitrogen’ is released through cell exudates and cell death/decay and becomes available to other photosynthetic organisms (Fig. 1.1). 
N₂ fixing microorganisms (Sec. 1.3.1) are a key component of biogeochemical cycles driving a significant amount of new production (480–960 Tg C yr⁻¹) in oligotrophic regions (Mahaffey et al. 2005). Globally 100-200 Tg of new nitrogen is introduced to the ocean system per year (Karl et al. 2002).

Box 2: NITROGEN FIXATION

The process of N₂ fixation (diazotrophy) is fundamental in the marine environment as it sustains primary production in oligotrophic regions of the ocean.

Through this process nitrogen gas (N₂) from the atmosphere is converted to ammonia (NH₃). The enzyme nitrogenase, an enzyme inactivated by oxygen (O₂), is required for the catalysis of this reaction. The stability of N₂ due to its triple bond renders N₂ fixation an energetically costly process:

\[ \text{N}_2 + 8\text{H}^+ + 16\text{ATP} + 8\text{e}^- \rightarrow 2\text{NH}_3 + \text{H}_2 + 16\text{ADP} + 16\text{P}_i \]

The identity of the single limiting nutrient controlling oceanic productivity over geological timescales has been the subject of a long scientific debate with P and N at the centre of it (Falkowski 1997; Tyrrell 1999). The ‘geochemists view’ is that as biological N₂ fixation can relieve organisms of external N requirements while there is no atmospheric source for P, supply of the latter will ultimately control the ocean’s productivity over time (Redfield 1958; Tyrrell 1999). On the other hand, the ‘biologists view’ states that due to loss of fixed N through denitrification and/or anammox reactions, and as N appears at a deficit compared to P, in the surface oligotrophic tropical/subtropical oceans, production over long timescales will be controlled by the process of N₂ fixation (Falkowski 1997; Falkowski et al. 1998).

In reality the theory of nutrient limitation can be applied to different levels, from the smallest cellular level where a nutrient limits the activity/proliferation of a single cell at a specific moment in time and particular oceanic location, to restriction of productivity across the world oceans over long time scales. The term ‘ultimate limiting nutrient’ was coined to describe the latter (Tyrrell 1999). Between these two levels, a nutrient can restrict the growth of a community at a specific time and place and this is defined as the ‘proximal limiting nutrient’. Therefore, P has been suggested to take the role of the ultimate limiting nutrient while N of the proximal limiting nutrient (Tyrrell 1999).
Figure 1.2: The link between nutrients and primary production in the ocean. (A) Concentrations of nitrate are depleted across the low latitude oligotrophic oceans and thus (B) limit primary production (chlorophyll concentrations). Chlorophyll concentration also remain relatively low in locations referred to as high nutrient-low chlorophyll (HNLC) regions which include the equatorial and subarctic Pacific Ocean and the Southern Ocean (figure from Gledhill & Buck 2012).

However, restriction of N₂ fixation through limitation of the micronutrient iron (Fe) complicates this justification (Martin 1990; Moore & Doney 2007; Moore et al. 2009; Martínez-Garcia et al. 2011) suggesting that this nutrient also has an important role in the control of oceanic primary production. In addition to that, Fe appears to have a direct control on productivity of large oceanic areas with high N and P concentrations such as the equatorial and subarctic Pacific Ocean and the Southern Ocean. These areas maintain low levels of primary production and are referred to as high nutrient-low chlorophyll locations (HNLC) (Fig. 1.2) (Martin et al. 1994; Boyd et al. 2000; Tsuda et al. 2003; Buesseler et al. 2004; Blain et al. 2007).
Limitation of marine nutrients P and Fe is briefly reviewed below taking into account speciation and complexation of these nutrients in the water as well as their biological importance and interactions with marine microorganisms (Fig. 1.3).

1.2.1 Phosphorus

Phosphorus is essential for the construction of phospholipid cell membranes and nucleotides used as the building blocks of (diox)ribonucleic acids (DNA and RNA) as well as nucleoside triphosphates (eg. ATP) that constitute the cell’s energy currency. Although it is the 11th most common element in the Earth’s crust it is mostly confined to the lithosphere from where weathering and riverine/aerial transfer delivers it to the oceans (Compton et al. 2000). Most (~75%) of the new P entering the ocean is in the form of particulate (> 0.2 μm) inorganic P (PIP) but some dissolved (< 0.2 μm) inorganic P (DIP) as well as particulate and dissolved organic P (POP and DOP) is also supplied (Compton et al. 2000). This P will be incorporated in growing cells (Fig. 1.3), travel through the food web and sink upon cell death, but a large proportion will be returned/remade available to the surface ocean largely as DIP through upwelling and lateral transfers (Mahaffey et al. 2004). Microbial transformations of P are numerous and support a complex cycle which is currently the subject of intensive research (reviewed by Karl 2014).

Characterisation of P forms, their concentrations and fluxes is challenging even when more than 50 analytical procedures are currently employed for their identification (Karl & Björkman 2001). Certain DOP/PIP molecules including most organophosphonates (Pn) (+3 valence state), reduced P compounds with a direct C-P bond which makes them more stable than organophosphates (+5 valence state, C-O-P bonds), resist hydrolysis which is used to quantify ‘total P’ leading to inaccuracies in the calculated concentrations. Another limitation to the methods is that organic P concentrations are calculated by subtracting phosphate (PO$_4^{3-}$) concentrations from the ‘total P’. This disregards reduced inorganic P sources phosphite, PO$_3^{3-}$ (+3 valence state), and hypophosphite, PO$_2^{3-}$ (+1 valence state) which might be found in significant concentrations in aquatic environments (Pasek et al. 2014). In addition, PO$_4^{3-}$ itself is hard to track at concentrations below 1nmol L$^{-1}$ and fast turnover times of below 1 hour sometimes encountered in oligotrophic regions (Karl 2014).

From the information currently available we can induce that PO$_4^{3-}$ concentrations vary across the surface oceans by more than 1000 fold and are heavily depleted at low latitudes. The DOP pool exceeds that of PO$_4^{3-}$ often by more than 10 times and can be comprised by
25% of Pn (Clark et al. 1998; Kolowith et al. 2001; Karl 2014). Also, a significant proportion (up to 25%) of the dissolved P ocean inventory could be present as reduced inorganic P, \( \text{PO}_3^{3-} \) and \( \text{PO}_2^{3-} \) (Pasek et al. 2014).

Bottle incubation experiments have previously suggested that the ocean is limited or colimited (in combination with other nutrients) by \( \text{PO}_4^{3-} \) at locations such as the Central to North Atlantic and Pacific Oceans (Wu et al. 2000; Sañudo-Wilhelmy et al. 2001; Mills et al. 2004; Moutin et al. 2005; Mills et al. 2008). For the Atlantic Ocean this has been explained by an Fe influx providing an excess of the nutrient to the North, and therefore inducing P drawdown to limiting concentrations by growing diazotrophs (Wu et al. 2000; Moore et al. 2009). In contrast, P is more abundant in the South Atlantic where diazotrophs are limited by Fe (Moore et al. 2009).

Figure 1.3: Interactions of phosphorus and iron forms with the biotic and abiotic environment in the ocean. Different forms of P (left) and Fe (right) in the ocean are interconverted due to physical and biological activities (blue arrows) and can be taken up/released by marine organisms (green arrows). Valence states of P and Fe forms are presented in brackets.
1.2.2 Iron

Iron is the fourth most common element in the earth’s crust and is delivered to the oceans through fluvial and aerial sources as well as hydrothermal vents. Aerial dust deposition from the major deserts is the predominant input to the open oceans (Moore & Braucher 2008). At circumneutral pH and the oxidising conditions encountered in the ocean, Fe$^{2+}$ is quickly oxidised to Fe$^{3+}$ which has a much lower solubility and is more likely to precipitate out of the surface waters. This leaves the oceans highly anemic.

One factor affecting the solubility of dust Fe is mineralogy. Iron, which can constitute 4-7% of the total desert particle mass, is classified as either ‘structural Fe’ found in the crystal lattice of aluminosilicate minerals (clays, quartz and feldspars) or ‘free Fe’ in (hydr)oxides (hematite, magnetite, goethite). Mineralogy of dust varies depending on its source but clays make up ~48-82% of the dust mass, and although their Fe content is low (~2-23%) compared to (hydr)oxides (~58-77%), they have the largest contribution (96%) to dissolved Fe because of its higher solubility (Journet et al. 2008). The solubility of dust Fe, shown to vary between 0.01 and 80% (Mahowald et al. 2005), will also depend on several processes both atmospheric (Photo(chemistry), physical processing, transport and mixing) and oceanic (dissolved Fe concentrations, Fe-binding ligands, biological processes, particle adsorption and the kinetics of dissolution) (Baker & Croot 2010).

The majority (99%) of oceanic dissolved Fe (< 0.2 μm) was previously suggested to be organically bound (Gledhill & Berg 1994; Rue & Bruland 1995; Wu & Luther 1995; Gledhill & Buck 2012) but recent evidence suggests that a considerable amount could be present as inorganic dissolved colloidal (0.02-0.2 μm) Fe (Fitzsimmons et al. 2015). The organic Fe-binding ligands currently remain largely uncharacterised (Gledhill & Buck 2012) but their significance is paramount as they contribute significantly in maintaining Fe in solution and making it more liable to photoreduction (Barbeau et al. 2001; Wagener et al. 2008).

Iron is a critical element for growth of phytoplankton whose photosynthetic apparatus require as many as 23-24 Fe atoms (Shi et al. 2007). At the same time diazotrophs require ~2.5-5.2 times more Fe than non-diazotrophs (Sañudo-Wilhelmy et al. 2001) as nitrogenase, the enzyme responsible for N$_2$ fixation, requires an additional 38 Fe atoms per monomer. Potentially 80-90% of the cellular Fe-quota is associated with nitrogenase in some diazotrophs (Richier et al. 2012).
The potential importance of Fe in constraining primary production in the ocean was first stressed by John H. Martin in his ‘Iron Hypothesis’. For more than 25 years his words: ‘Give me half a tanker of iron and I will give you an ice age’ stimulated the scientific community (Martin 1990). Multiple large scale Fe release experiments, as well as shipboard bottle incubation experiments were designed to test Martin’s Hypothesis and confirmed the importance of Fe in HNLC locations (Boyd et al. 2007). In addition to these locations, growth of diazotrophs in various oligotrophic regions appears to be limited by Fe (Moore et al. 2009). More recently, the Iron Hypothesis was supported by data from a Subantarctic Atlantic sediment core indicating a connection between global cooling over geological time scales, productivity and desert dust fluxes to the ocean (Martinez-Garcia et al. 2014).

1.2.3 Paradigms of nutrient limitation

The first conceptual model used to describe nutrient limitation in the ocean was based on work by J. von Liebig who indicated how the scarcest nutrient can control the levels of production (Leibig’s Law of the minimum) (de Baar 1994), and F. F. Blackman who extended this theory to the effect of a limiting factor on constraining photosynthesis and growth rates (Blackman’s Law of limiting factors) (Blackman 1903).

More recent observations, that often additions of more than one nutrients can stimulate growth of phytoplankton, have introduced the concept of ocean ‘colimitation’ (Saito et al. 2008). This describes the situation whereby more than one nutrient simultaneously limit production. There are different scenarios through which colimitation can occur (Saito et al. 2008):

- Independent nutrient colimitation: The level of two nutrients can be reduced to equally minimal levels so as additions of both induce growth (Saito et al. 2008).
- Biochemically dependent colimitation: One nutrient limits the ability to acquire a different nutrient. An example of this is provided by new evidence indicating that Fe is a cofactor of the P utilisation enzyme alkaline phosphatase which is consequently controlled by the availability of both nutrients (Yong et al. 2014; Mahaffey et al. 2014).
- Biochemical substitution colimitation: Two nutrients can substitute each other in specific biochemical functions. An example is the occurrence of superoxide dismutase SOD proteins (involved in cellular defence against oxidative stress) with different element binding properties (Ni, Mn, Fe or Cu and Zn) (Wolfe-Simon et al. 2005).
1.3 Marine Diazotrophy

$\text{N}_2$ fixation, essential in maintaining primary production under $\text{N}$-limitation, is highly interlinked with the P and Fe biogeochemical cycles (Falkowski 1997; Moore et al. 2009). This necessitates a better understanding of the identity and ecological properties of diazotrophic species.

1.3.1 The players

The characterisation of contributors to oceanic $\text{N}_2$ fixation and their distribution is still underway. To this end, molecular techniques for the identification of the diversity and distribution of the conserved gene $\text{nifH}$, which codes for a subunit of the catalytic enzyme of $\text{N}_2$ fixation, nitrogenase, have been utilised for the past 2 decades (Zehr et al. 1998).

Diazotrophic capabilities are encountered in both cyanobacterial and non-cyanobacterial species. Heterotrophic diazotrophs, identified through $\text{nifH}$ molecular analysis, appear to be widespread and at certain locations the dominant $\text{N}_2$ fixers (Farnelid et al. 2011). Their prevalence was indicated in environments such as the oxygen minimum zones and deep waters where photosynthetic organisms are not active (Halm et al. 2009; Fernandez et al. 2011; Farnelid et al. 2013). Diazotrophic cyanobacteria on the other hand, are found to dominate across the tropical and subtropical surface oceans (Fig. 1.4) (Stal 2009). They include filamentous species found between 26.5 to 30 °C and unicellular which can also exist in deeper, colder waters ranging between 15 and 30 °C (Moisander et al. 2010).

**Unicellular cyanobacteria (UCYN)**

These small (< 10 μm) diazotrophs (Goebel et al. 2007) were first identified at station ALOHA using the $\text{nifH}$ technique (Zehr et al. 1998) and can be categorised in three groups: UCYN-A, UCYN-B, UCYN-C (Table 1A). Until recently, they have been generally overlooked but recent studies suggest a higher contribution towards the total diazotrophic community than previously assumed (Montoya et al. 2004). They are currently thought to fix and release $\text{N}$ at rates similar to the more well studied diazotrophs (i.e. *Trichodesmium*) (Falcon et al. 2004; Berthelot et al. 2015, Martinez-Pérez et al. 2016) particularly in winter when conditions are not optimal for the latter (Goebel et al. 2007; Lee Chen et al. 2014).

The UCYN-A *Candidatus Atelocyanobacterium thalassa* have a highly streamlined genome lacking photosystem II (PSII) complexes and the ability to fix $\text{CO}_2$ (Zehr et al. 2008; Tripp et al. 2010; Bombar et al. 2014). They form symbiosis to the single celled Prymnesiophyceae
algae (Thompson et al. 2012; Hagino et al. 2013) with which they exchange 85-95% of their fixed N for carbon (Thompson et al. 2012; Krupke et al. 2015). New research indicates that different UCYN-A clades (UCYN-A1 and UCYN-A2) display ‘partner fidelity’ with their algae host (Thompson et al. 2014; Cornejo-Castillo et al. 2016). The cultivated group UCYN-B *Crocosphaera watsonii* found widespread in the ocean (Mazard et al. 2004; Hewson et al. 2009; Webb et al. 2009), can form symbiotic relationships with diatoms (Carpenter & Janson 2000; Foster et al. 2011) but is mainly free living and can form colonial aggregations (Foster et al. 2013). UCYN-C contains the free-living cultivated species *Cyanothece* (Reddy et al. 1993) and TW3 (Taniuchi et al. 2012) and the importance of this group was suggested in the recent VAHINE mesocosm experiment at the New Cledonian lagoon (Bonnet et al. 2016; Hunt et al. 2016).

**Figure 1.4: Marine diazotrophy in nutrient limited oceans.** As nitrate by definition cannot limit diazotrophs, N$_2$ fixation thrives at low latitudes (rates indicated as fill colours), with *Trichodesmium* found across the majority of these regions. Other nutrients including iron (green outline), which is primarily supplied through Aeolian dust deposition (arrows), and phosphorus (yellow outline) are limiting diazotrophic growth (Sohm et al. 2011).
Filamentous nitrogen fixing cyanobacteria

DDAs: Diatom Diazotroph Associations (DDAs) are composed of a diatom host and a diazotroph symbiont (Richelia and Calothrix lineages) (Table 1A) (Janson et al. 1999; Foster & Zehr 2006). They are hypothesised to be important contributors to organic carbon export from the surface ocean due to the heavy, high silicon containing diatom cell walls (Subramaniam et al. 2008).

The interactions between the host and diazotroph are intricate and presently not fully understood but recent data suggests that DDA formation is stimulated by low nitrate availability (Tuo et al. 2014). Symbiosis can be obligate or facultative (Peters & Meeks 1989; Kneip et al. 2007; Ran et al. 2010; Hilton et al. 2013) and the arrangement of the relationship varies: Richelia is found inside Rhizosolenia and Hemiaulus cells (Carpenter et al. 1992) while Calothrix is extracellular to Chaetoceros spp. and can be cultured independently (Foster et al. 2010). Surprising new evidence also identified an association of Calothrix to the abundant free living colonial diazotroph Trichodesmium (Momper et al. 2015). Different life strategies are controlled by- and have shaped- the genomes of the two diazotroph symbionts Richelia and Calothrix. The former possesses a streamlined genome compared to the latter, with genes involved in N metabolism lacking. This mechanism possibly ensures stability of N₂ fixation as the process is essential for the symbiotic partnership (Hilton et al. 2013).

Table 1A: Characteristics of the major diazotrophic cyanobacteria groups.

<table>
<thead>
<tr>
<th>Diazotroph group</th>
<th>Morphology</th>
<th>Strategy of N₂ fixation</th>
<th>Lifestyle</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCYN-A</td>
<td>Unicellular</td>
<td>Spatial (host fixes C)</td>
<td>Symbiotic (host: Prymnesiophyceae)</td>
</tr>
<tr>
<td>Candidatus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atelocyanobacterium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>thalassa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UCYN-B</td>
<td>Unicellular</td>
<td>Temporal (N₂ fixation at</td>
<td>Free-living</td>
</tr>
<tr>
<td>Crocosphaera</td>
<td></td>
<td>night)</td>
<td></td>
</tr>
<tr>
<td>watsonii</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UCYN-C</td>
<td>Unicellular</td>
<td>Temporal (N₂ fixation at</td>
<td>Free living</td>
</tr>
<tr>
<td>Cyanothece spp</td>
<td></td>
<td>night)</td>
<td></td>
</tr>
<tr>
<td>DDA</td>
<td>Filamentous</td>
<td>Spatial (form heterocysts)</td>
<td>Symbiotic (host: diatoms)</td>
</tr>
<tr>
<td>Richelia spp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calothrix spp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trichodesmium spp.</td>
<td>Filamentous/</td>
<td>Combinatory (form diazocytes and fix at</td>
<td>Free-living</td>
</tr>
<tr>
<td></td>
<td>Colonial</td>
<td>midday)</td>
<td></td>
</tr>
</tbody>
</table>

Epibionts
Similar to UCYN-A the diazotroph symbiont of DDAs transfers fixed N to the host which in the case of Richelia can amount to approximately 97% of the N\textsubscript{2} fixed (Foster et al. 2011). The benefit to the diazotroph symbionts is not currently clear as, unlike UCYN-A, they are fully capable of photosynthetically fixing their own carbon. It is possible, that transfer of additional carbon, or other nutrients, from the host to the diazotrophs, makes the symbiosis advantageous (Krupke et al. 2015).

**Trichodesmium:** *Trichodesmium* is widely distributed in an area covering approximately half of the earth’s surface across the tropical and subtropical oceans (Fig. 1.4) and accounts for ~50% of the total marine N\textsubscript{2} fixation (Capone et al. 1997; Westberry & Siegel 2006). Two of the most important species are *T.erythraeum* and *T.thiebautii* with the former (isolated in 1991) (Prufert-Bebout et al. 1993) being the most well studied strain of this species.

*Trichodesmium* is polymorphic, found in filaments of ~100-200 cells that come together to form puff (Fig. 1.5Bi), tuft (Fig. 1.5Bii), and bowtie colonies (Table 1A). The colonies are referred to as consortia due to their function as microhabitats for an array of microorganisms clustered (at high densities compared to the surrounding water) in and around the colonies (Sheridan et al. 2002). These microorganisms include a diverse community of other cyanobacteria, heterotrophic bacteria, eukaryotes and phages (Hewson et al. 2009) which are distinct to the community of the surrounding environment (Hmelo et al. 2012). In addition, *Trichodesmium* colonies act as biological activity hotspots where epibionts potentially benefit from the N and carbon released from the *Trichodesmium* cells. Interactions within the consortia community appear to be central to the controls of nutrient utilisation, with quorum sensing shown to stimulate the expression of genes involved in P utilisation (Van Mooy et al. 2012).

First reports of *Trichodesmium* were recorded by Captain James Cook and Sir Joseph Banks on board HMS *Endeavour* on her first voyage (Banks 1770) and the organism was first described by Ehrenberg (1830) (Box 3). It is thought that the Red Sea was named from the colouration of the water during *Trichodesmium* blooms. These blooms are visible from space and remote sensing can now report they extend over distances of $\sim 5.4 \times 10^{6}$ km\textsuperscript{2} (Westberry & Siegel 2006).
Figure 1.5: *Trichodesmium* morphology. (A) Copepods graze on *Trichodesmium* colonies in the environment (image taken during the PABST cruise in the Sargasso Sea). (B) *Trichodesmium* filaments come together to form colonies (puff (i) and tuft (ii) morphotype presented). (C) Lugol’s stained *Trichodesmium* filaments visualised under light microscopy. Diazocytes (d) are recognised by their lighter, green colouration due to reduced carbohydrate storage compared
Trichodesmium blooms, which develop rapidly likely in response to nutrient influxes, significantly enhance N inputs to the surrounding environment while sequestering large amounts of atmospheric carbon (Capone et al. 1997; Lenes et al. 2001; Westberry & Siegel 2006; Luo et al. 2012). These blooms are observed to collapse rapidly (Rodier & Le Borgne 2008; Rodier & Le Borgne 2010; Bergman et al. 2013; Spungin et al. 2016) possibly due to viral activity (Hewson et al. 2004) or autcatalytic programmed cell death (PCD) likely stimulated by nutrient stress (Berman-Frank et al. 2004; Berman-Frank et al. 2007; Bar-Zeev et al. 2013; Spungin et al. 2016). Recent observations from the VAHINE mesocosm experiments indicate no significant changes in viral abundance but enhanced caspase activity during the Trichodesmium bloom demise in support of PCD being the mechanism of bloom termination (Spungin et al. 2016).

**Box 3: First references to Trichodesmium blooms**

“On December 10th I observed there the astonishing phenomenon of the bloody coloration of the entire bay which forms the harbour of Tor. The open sea outside of the coral reef which shunts off the harbour was colourless as usual. The short waves of the calm sea conveyed to shore a slimy mass which appeared blood-red in the sunlight and was deposited in the sand of the beach, so that at ebb tide the entire bay, which takes a good half hour to cross, was given a blood red border.”

*(Ehrenberg 1830)* as referenced by (Carpenter et al. 1992)

“On the 15th of July (1843), the blazing sun of Arabia woke me abruptly, shining suddenly on the horizon without any dawn and in all its splendour. I leaned my elbows rather mechanically on a windowsill at the stern of the ship to seek a vestige of fresh night air before the heat of the day devoured it. It was my surprise to see the ocean astern of the ship coloured red as far as the eye could reach! I ran up on the deck and saw the same phenomenon on all sides.”

*Letter by E.Dupont in 1843* (Wille 1904) as referenced by (Carpenter et al. 1992)

“We sailed from Bahia. A few days afterwards, when not far distant from the Abrolhos Islets, my attention was called to a reddish-brown appearance in the sea. The whole surface of the water, as it appeared under a weak lens, seemed as if covered by chopped bits of hay... ...Their numbers must be infinite: the ship passed through several bands of them, one of which was about ten yards wide, and, judging from the mud-like colour of the water, at least two and a half miles long.”

*(Darwin 1909)* The Voyage of the Beagle
Despite *Trichodesmium* rarely being found in sediments traps (Walsby 1992; Scharek et al. 1999; Chen et al. 2003), a strong downward pulse of organic matter has been suggested to follow PCD indicating a possible link of this process to the efficiency of the carbon pump (Bar-Zeev et al. 2013). In addition, recent evidence indicated export of carbon though non-diazotrophs (i.e. diatoms) that benefit from fertilisation of the surrounding waters, following *Trichodesmium* blooms (Bonnet et al. 2015). Indeed, *Trichodesmium* was reported to transfer twice as much fixed N\(_2\) to non-diazotrophs (both heterotrophs and autotrophs) compared to unicellular diazotrophs (Bonnet et al. 2015).

### 1.3.2 Oxygen avoidance

The ability of diazotrophic cyanobacteria to fix N\(_2\) clashes with the process of photosynthesis as O\(_2\) produced during the latter is detrimental to nitrogenase. Therefore, mechanisms have evolved to segregate the two processes spatially, or temporally (Berman-Frank et al. 2003).

Spatial segregation is encountered in filamentous cyanobacteria which can form thick-walled, PSII-lacking cells referred to as heterocysts to accommodate N\(_2\) fixation (Fig. 1.6). Both symbiotic diazotrophs like *Richelia* (Gomez et al. 2005) (Table 1A) and free-living like the freshwater model-cyanobacteria *Anabaena* (Kumar et al. 2010) employ this strategy. Unicellular diazotrophs on the other hand, do not have the luxury for division of labour between cells and therefore temporally segregate photosynthesis and N\(_2\) fixation instead (Fig. 1.6, Table 1A). For example, *Crocosphaera* is a nocturnal N\(_2\) fixer with a protein diel cycle that involves *de novo* synthesis of the NifH at the start of the scotoperiod when photosynthesis in not functioning (Saito et al. 2011). The same strategy is followed by *Cyanothece* (Toepel et al. 2008).

Unique amongst other diazotrophs, *Trichodesmium* employs a less strict variation of both modes of segregation simultaneously (Fig. 1.6, Table 1A). Although not heterocyst forming, ~15 % of each filament develops specialised nitrogenase containing (Berman-Frank et al. 2001) (but not terminally differentiated) cells with degraded gas vacuoles and glycogen granules (Sandh et al. 2012) arranged in zones of ~2-30 cells (Bergman et al. 2013) (Fig. 1.5C). At the same time, and although both photosynthesis and N\(_2\) fixation occur during the photoperiod, it possesses its own circadian clock synchronising a midday downregulation of photosynthesis (low or negative O\(_2\) evolution) (Berman-Frank et al. 2001) with enhanced O\(_2\) scavenging mechanisms (respiration and Mehler reaction) (Kranz et al. 2009) and a peak in nitrogenase activity (Berman-Frank et al. 2001).
Trichodesmium’s fixing strategy might be advantageous by allowing direct transfer of energy from photosynthesis to fuel N\textsubscript{2} fixation contrary to when the two processes are separated in time and thus utilisation of energy reserves is required (Staal et al. 2007). The flexibility of the non-terminally differentiated diazocytes as opposed to heterocysts could also be a benefit to survival in the rapidly changing oceanic conditions this organism encounters. It should be added, that an understanding of the controls on diazocyte formation and their mechanism of action could also be biotechnologically interesting in the field of algal biofuels as they could provide genetic engineering strategies coordinating aerobic and anaerobic chemical reactions.

Figure 1.6: Strategies of N\textsubscript{2} fixation-photosynthesis segregation. Different groups of diazotrophs employ different approaches: (A) During temporal separation (e.g. in Cyanothecae) N\textsubscript{2} fixation is carried out during the scotoperiod, (B) spatial separation (e.g. in Anabaena) involves specialised N\textsubscript{2} fixing cells (heterocysts), (C) while a combinatory strategy (in Trichodesmium) is observed when specialised but not terminally differentiated N\textsubscript{2} fixing cells (diazocytes) have their peak activity at midday when photosynthesis is depressed (figure from Snow 2014).
Despite all the advantages it confers *Trichodesmium*’s synchronous N\textsubscript{2} and CO\textsubscript{2} fixation requires the maintenance of a higher cellular Fe quota (~66-87 μmol mol\textsuperscript{-1} carbon) compared to nocturnal N\textsubscript{2} fixers like *Crocosphaera* (~16 μmol mol\textsuperscript{-1} carbon) (Tuit et al. 2004) due to high Fe demands of both the photosynthetic and diazotrophic lifestyles (Geider & La Roche 1994; Shi et al. 2007).

### 1.4 Living with less- lessons from marine microorganisms

It has been previously discussed (Sec. 1.2) that P and Fe are key nutrients controlling the primary productivity of the oceans. Phytoplankton and amongst them diazotrophs whose ecological niche spreads across P limited oligotrophic gyres (Fig. 1.4) and their N\textsubscript{2} fixing qualities come at a cost of Fe, are forced to cope with the increased pressures. They therefore display a series of highly streamlined adaptations that ensure efficient use of the available nutrients. With focus on the globally important diazotroph *Trichodesmium* some of the adaptations of marine producers to low environmental P and Fe are reviewed below:

#### 1.4.1 Cellular strategies to cope with P stress

**Regulation:** The regulatory mechanism coordinating the cell’s response to low P is controlled by the Pho regulon first identified in *E.coli* (Wanner & Chang 1987). This functions through a two component system involving an inner membrane histidine kinase sensor (SphS/PhoR), and a transcriptional response regulator (SphR/PhoB). When P is unavailable to the transmembrane subunit it phosphorylates and activates the regulator which interacts with DNA sequences referred to as PHO boxes. Interestingly, even though *Trichodesmium* seems to be lacking the transmembrane SphS/PhoR component, the transcriptional regulator SphR/PhoB appears to regulate its PO\textsubscript{4}\textsuperscript{3-}-deficiency response by inducing the transcription of transporters and enzymes for PO\textsubscript{4}\textsuperscript{3-}, organophosphate, and Pn utilisation (Dyhrman et al. 2006; Su et al. 2007; Orchard et al. 2009).

**Conservation of resources:** A strategy employed during times of nutrient scarcity is the management of intracellular reserves to maintain cellular activity at the lowest possible nutrient expense. *Trichodesmium* cells growing with reduced concentrations of P have a lower P quota indicating that mechanisms of lowering the cellular requirements for the nutrient are employed by this organism (White et al. 2006a, 2010).
An example of such a mechanism is the substitution of P containing, phosphatidylglycerol (PG), with non-phosphorus containing, sulfoquinovosyldiacylglycerol (SQDG), membrane lipids under P-deficient conditions. This was characterised in laboratory cultures of *Trichodesmium* as well as for other major marine primary producers like *Synechococcus*, *Prochlorococcus*, the diazotroph *Crocosphaera*. (Van Mooy et al. 2009) and diatoms (Martin et al. 2011). Observations that the low P Sargasso Sea is characterised by a higher abundance of non-P lipids, lowering the P quota of organisms’ lipid membranes compared to the South Pacific (Van Mooy et al. 2008; Van Mooy et al. 2009) reaffirm the significance of this strategy *in situ*.

**Storage:** Storage of cellular P as inorganic polymer polyphosphate (polyP) is thought to create luxury reserves of the nutrient in P-replete conditions for use during nutrient stress (Rhee 1973; Ault-Riché et al. 1998; Werner et al. 2005). However, *Trichodesmium* and other phytoplankton have been indicated to increase polyP concentrations in P-deficient conditions (Orchard et al. 2010a; Martin et al. 2014) for reasons not currently characterised.

**Vertical migration:** Diurnal vertical migration is a mechanism used by phytoplankton including diatoms and flagellates to access nutrients below the nutricline when they are depleted in surface waters (Eppley et al. 1968; Hall & Pearl 2011; Villareal et al. 2014). It is suggested that *Trichodesmium* uses cellular gas vesicles to regulate its buoyancy and acquire P at depth during the night, before rising to the surface to resume photosynthetic activity (Karl et al. 1992; Villareal & Carpenter 2003; White et al. 2006a).

**P utilisation:** As in many other organisms, PO$_4^{3-}$ uptake by *Trichodesmium* appears to be mediated by the high affinity P specific transporter (Pst). The organism possesses homologs to the ABC transporter (PstABC) composed of the inner membrane components PtsA and PstC, the ATPase PstB and the PO$_4^{3-}$ binding proteins PstS/SphX (Orchard et al. 2009) (Fig. 1.7). Interestingly, *sphX* was previously identified to be regulated by PO$_4^{3-}$ concentrations while *pstS* was not. The two genes are colocalised in the genome but *sphX* is separated to *pstS* and the transporter genes *pstABC* by a 19 545 bp gap suggesting regulation by separate promoters (Su et al. 2007; Orchard et al. 2009). No homologues of the low affinity PO$_4^{3-}$ transporter Pit (Fig. 1.6) have been identified in this species (Willsky et al. 1973).

Utilisation of organophosphorus (OP) is facilitated by enzymes that catalyse release of PO$_4^{3-}$ which can consequently be internalised by the Pst transporter. In the case of
organophosphate these enzymes are alkaline phosphatases (APs) encoded by genes phoX and phoA (Fig. 1.7). *Trichodesmium*, which relative to the microbial community appears to be a better competitor for use of organophosphate compared to PO$_4^{3-}$ (Orchard et al. 2010b), has one putative homologue of phoA (preceded by a Pho-box related promoter sequence) and two of phoX (only phoX1 has Pho-box related promoter sequence). The expression of phoA and phoX1 is upregulated under P limitation and AP activity was previously used to assess the organism’s P state *in situ* (Dyhrman et al. 2002; Webb et al. 2007; Orchard et al. 2009).

APases can be found in the periplasm of marine bacteria but also intracellularly (Fig. 1.7) to facilitate organophosphate catabolism after direct uptake through Upg transporters (Luo et al. 2009). These transporters, first characterised in *E. coli*, are composed of the periplasmic binding UgpB, the inner membrane UgpAE and the ATPase UgpC (Overduin et al. 1988; Brzoska & Boos 1988) (Fig. 1.7), and appear to be abundant in the marine environment (Luo et al. 2009). No obvious homologous exist in *Trichodesmium*.

AP activity (APA) is a commonly used biomarker of P-limitation *in situ* (Sebastián et al. 2004a, 2004b; Dyhrman & Ruttenberg 2006; Sato et al. 2013), however, it only provides an overall community P-stress index, which due to taxonomic heterogeneity might not be representative of the efficiency of processes affecting biogeochemical cycles (Lomas et al. 2004). In addition, recent evidence suggests an even more critical pitfall of this technique as PhoX appears to require Fe for its production (Yong et al. 2014). This is problematic if AP activity is used to assess P limitation in a highly Fe variable environment (Mahaffey et al. 2014).

The predominant route for Pn catabolism is through the enzyme C-P lyase (Fig. 1.7) which acts to hydrolyse the C-P bond of a wide variety of substrates. Steps in better understanding the mode of the dephosphonation reaction have recently been made but the pathway is still not fully characterised (Kamat et al. 2011; Seweryn et al. 2015). The C-P lyase core components are encoded by genes *phnGHJKLM* found in a gene cluster together with Pn ABC transporter genes *phnCDE*. In *E. coli* but not *Trichodesmium*, *phnF* and *phnNOP* thought to act as a transcriptional repressor and regulator respectively (Hove-Jensen et al. 2010, 2011), are also members of the C-P lyase cluster. The expression of the *phn* cluster is under the control of PO$_4^{3-}$ concentrations in *Trichodesmium* and appears to be coordinated by the Pho regulon (Dyhrman et al. 2006; Su et al. 2007).
Figure 1.7: The Phosphorus and Iron uptake and utilisation pathways in microorganisms. Mechanisms of phosphate ($\text{PO}_4^{3-}$), phosphite ($\text{PO}_3^{3-}$), hypophosphite ($\text{PO}_3^{3-}$), organophosphate and organophosphonates (Pn) have been previously indicated to be employed for P utilisation. Similarly, phytoplankton is known to be able to utilise dust, organically bound iron and both reduced ($\text{Fe}^{2+}$) and oxidised ($\text{Fe}^{3+}$) free inorganic iron ($\text{Fe}'$) by employing a broad spectrum of uptake pathways.
In addition to the reduced organic form, P can possibly also be found in the ocean in its reduced inorganic state as $\text{PO}_3^{3-}$ and $\text{PO}_2^{3-}$ (Fig. 1.3). The potential prevalence of these sources in aquatic environments (Pasek et al. 2014) and their importance to *Trichodesmium* colonies which seem capable of both producing and possibly taking up $\text{PO}_3^{3-}$ (Van Mooy et al. 2015) are currently not well characterised. Gene clusters for $\text{PO}_3^{3-}$ (*ptx*) and $\text{PO}_2^{3-}$ (*htx*) utilisation, first characterised in *Pseudomonas stutzeri*, are members of the Pho regulon in this organism. They include genes encoding ABC transporters (*ptxABC* and *htxBCD*) and enzymes NAD:phosphite oxidoreductase (*ptxD*) and hypophosphite dioxygenase (*htxA*) required for $\text{PO}_3^{3-}$ and $\text{PO}_2^{3-}$ oxidation respectively (Metcalf & Wolfe 1998; White & Metcalf 2004) (Fig. 1.7). It is worth noting that both of these reactions can potentially be carried out by C-P lyases and APases as well, although the reaction through PtxD is more energetically favourable.

Homologues of the *ptx* operon have recently been identified in marine cyanobacteria and diazotrophs, including *Trichodesmium*, and have been characterised to be responsible for $\text{PO}_3^{3-}$ utilisation in *Prochlorococcus* MIT930 (Martínez et al. 2012). On the other hand $\text{PO}_3^{3-}$ appears to be inaccessible or have a deleterious effect on microalgal species *Chlamydomonas reinhardtii*, *Botryococcus braunii* and *Ettlia oleoabundans* (Loera-Quezada et al. 2015). The ability of *Trichodesmium* to utilise $\text{PO}_3^{3-}$ and the role of the *ptxABCD* homologous gene is further examined in Chapter 3 of this work.

1.4.2 Cellular strategies to cope with Fe stress

**Regulation:** The cellular response to Fe-limitation is coordinated by the ferric uptake regulator (Fur) which was first discovered in *E. coli* (Hantke 1981; Bagg & Neilands 1987a, 1987b). Fur binds to both DNA and Fe (Bagg & Neilands 1987a; Deng et al. 2015) and acts as a transcriptional repressor when the metal is abundant. Under Fe-deficiency it dissociates from DNA to allow activation of genes involved in the organisms’ stress response (Lee & Helmann 2007; Osman & Cavet 2010). Despite its role as a repressor, Fur can also act as an activator directly or through interactions with regulatory small RNAs (sRNAs) (Delany et al. 2002; Massé & Gottesman 2002).

**Conservation of resources:** To reduce their cellular Fe demand marine microorganisms are known to replace Fe with non-Fe cofactors in certain enzymes and electron transporters. Examples are the electron transporters: ferredoxin which is replaced by flavodoxin
Cells are also able to reduce Fe demands by regulating expression of multisubunit complexes. For example, under Fe-deficiency the ratio of photosystem I (PSI) to photosystem II (PSII) drops (Fraser et al. 2013; Snow et al. 2015). This preferential reduction of one photosystem compared to the other is a result of their differential demand for Fe: PSI binds 12 Fe atoms per monomer and is thus much more Fe demanding compared to PSII which binds 3 (Shi et al. 2007). To compensate for the reduction of PSI:PSII the PSI antenna protein IsiA is stimulated (Bibby et al. 2001; Shi et al. 2007; Richier et al. 2012; Ryan-Keogh et al. 2012; Fraser et al. 2013; Snow et al. 2015) resulting in an increase of PSI light harvesting efficiency by up to seven times (Yeremenko et al. 2004).

Low-Fe-stimulated replacement proteins and the genes that encode them have been previously used as biomarkers for detecting Fe-limitation in the field (Erdner & Anderson 1999; Lindell & Post 2001; Chappell et al. 2012; Saito et al. 2014). This is a valuable technique as the bioavailable pools of Fe are presently not clearly characterised. By circumventing or complementing Fe concentration measurements the use of biomarkers allows direct sensing of the organisms’ perception of availability of the nutrient. However, care must be applied in interpreting the results as this proxy of Fe-limitation is not a direct measure of the N₂ and carbon fixation efficiencies.

**Storage:** Storage proteins are used to keep cellular Fe reserves for periods of low Fe supply. For example, ferritins are Fe storage proteins, which are widespread amongst bacteria, fungi, as well as plant and animal species, and form spherical complexes made of 24 subunits that can store 4500 Fe atoms (Andrews 1998; 2010). The related bacterioferritins, are only found in prokaryotes, and can sequester 2000 Fe atoms in their core. (Andrews 1998; 2010). They differ to ferritins in that they contain 12 hemes likely involved in Fe release (Yasmin et al. 2011). A third smaller protein complex from the same superfamily, termed ‘DNA-binding protein from starved cells’ or Dps, is found in prokaryotes and is made of 12 subunits that can store 500 Fe atoms (Castruita et al. 2006).

Studies have indicated the significance of Fe storage proteins in diatoms (Marchetti et al. 2009) and cyanobacteria for which bacterioferritins can associate with up to half of the cell’s Fe reserve and are vital for survival under Fe stress (Keren et al. 2004). Recently a diel regulation of ferritins has been demonstrated in phytoplankton. This is possibly associated...
with the replacement of damaged with de novo synthesised photosynthetic proteins during the dark period when they are not in use (Botebol et al. 2015). Securing released Fe in storage proteins during this process could be particularly beneficial during periods of Fe stress when external available pools are scarce and oxidative damage of photosynthetic machinery is augmented.

**Fe utilisation:** Highly efficient uptake mechanisms in cyanobacteria provide them with the ability to access various forms of Fe. Known accessible sources to *Trichodesmium* are presented in Table 1B. The uptake pathways that make these sources available are currently poorly characterized at the molecular level although such knowledge is important for understanding phytoplankton Fe homeostasis and its interaction with biogeochemical cycles.

The organic ligand pool of the surface oceans includes siderophores small chelating agents produced by bacteria and exported from the cell to bind Fe. Specialised transporters (TonB dependent transporters), subsequently take up the siderophore bound Fe (Neilands 1995). *Synechococcus* is the only marine cyanobacterial species identified to produce siderophores in addition to several marine heterotrophic bacteria (Ito & Butler 2005; Hopkinson & Morel 2009; Boiteau & Repeta 2015). Despite this, several phytoplankton species, and amongst them *Trichodesmium*, can engage in ‘siderophore piracy’ by utilising Fe from foreign siderophores (Achilles et al. 2003; Roe et al. 2012).

TonB-dependent transporters (TBDT) are composed of one or more outer membrane proteins utilising energy transferred by the periplasm-spanning TonB anchored on the plasma membrane. The energy transduction pathway starts with proteins ExbB and ExbD (forms homodimers with both ExbB and TonB in *vivo*) located in the plasma membrane (Ollis et al. 2009) where they utilise the proton motive force to induce conformational changes to TonB (Larsen et al. 1999; Brinkman & Larsen 2008; Ollis & Postle 2012; Ollis et al. 2012) (Fig. 1.7). The exact mechanism of energy transfer has not yet been elucidated. Although TBDTs involved in nickel and cobalt (Schauer et al. 2008), disaccharide and polysaccharide as well as vitamin B12 uptake have been identified, their major substrate is sidereophores and heme (Hopkinson & Barbeau 2012). Recent evidence suggests that ExbBD in *Synechocystis* sp. PCC 6803 (*Synechocystis* hereafter) are also involved in inorganic Fe uptake (Jiang et al. 2015). *Trichodesmium* possesses homologues of ExbBD and TonB but does not appear to have any obvious TBDTs (Hopkinson & Morel 2009).
The expression of TonB-transporter related genes is under the control of the Fur regulator in cyanobacteria (González et al. 2010, 2012) and differential expression of components was previously identified under Fe deficient conditions (Singh et al. 2003; Stevanovic et al. 2012). Furthermore, mutants of tonB and exbBD in Synechocystis display reduced growth compared to the WT under Fe-limitation (Jiang et al. 2012).

Siderophores transferred to the periplasm are carried across the inner membrane through ABC transporters Fhu and Fec, responsible for hydroxamate and citrate-type siderophore transport respectively (Krewulak & Vogel 2008; Stevanovic 2015). In cyanobacteria under reduced Fe conditions, components of this system are differentially regulated and growth of their deletion mutants is compromised (Stevanovic et al. 2012; 2013). Their transcriptional regulation was recently suggested to be controlled by Fur (González et al. 2014). Of the Fec/Fhu system subunits, *Trichodesmium* possesses only one protein with homology to the periplasmic hydroxamate binding FhuD (Chappell & Webb 2010).

### Table 1B: Sources of iron accessible to *Trichodesmium*.

<table>
<thead>
<tr>
<th>Fe form</th>
<th>Accessibility</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inorganic Fe</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fe(II)</td>
<td>yes</td>
<td>Roe et al. 2012</td>
</tr>
<tr>
<td>Fe(III)</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td><strong>Organically bound</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhodotorullic acid</td>
<td>yes</td>
<td>Achilles et al. 2003</td>
</tr>
<tr>
<td>(siderophore)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aerobactin (siderophore)</td>
<td>minimal</td>
<td>Roe et al. 2012</td>
</tr>
<tr>
<td>Desferrioxamine B</td>
<td>no/minimal</td>
<td>Roe et al. 2012</td>
</tr>
<tr>
<td>(siderophore)</td>
<td></td>
<td>Achilles et al. 2003</td>
</tr>
<tr>
<td>Protoporphyrin IX</td>
<td>minimal</td>
<td>Achilles et al. 2003</td>
</tr>
<tr>
<td>Citrate</td>
<td>yes</td>
<td>Roe et al. 2012</td>
</tr>
<tr>
<td>Dust</td>
<td>yes</td>
<td>Rubin et al. 2011</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Langlois et al. 2012</td>
</tr>
</tbody>
</table>

For transfer of uncomplexed free Fe\(^{3+}\) across the inner membrane bacteria utilise the ferric ion binding protein (Fbp) system/ferric uptake transport system (Fut) identified in pathogenic bacteria and *Synechocystis* respectively. It is composed of a periplasmic Fe-binding component (FbpA/FutA) a transmembrane (FbpB/FutB) and an ATPase (FbpC/FutC) subunit (Katoh et al. 2001a; Krewulak & Vogel 2008) (Fig. 1.7), homologues of which are found in *Trichodesmium*. Interestingly, *Synechocystis* possesses two paralogues of the periplasmic binding protein FutA. FutA2 appears to be periplasmic and complement FutBC
in Fe transport while FutA1 was indicated to be intracellular and have a function associated with protection of PSII under Fe-deficiency induced oxidative stress (Fulda et al. 2000; Tölle et al. 2002; Waldron et al. 2007). *Trichodesmium* has only one homologue, the identity of which is examined in Chapter 4.

For reduced Fe$^{2+}$, internalisation occurs through the Feo uptake system, which was first characterised in *E. coli*, and is composed of the inner membrane ATP-powered transporter FeoB, the transcriptional regulator FeoC and the protein of unknown function FeoA (Kammler et al. 1993; Cartron et al. 2006; Lau et al. 2013) (Fig. 1.7). Unlike *feoA*, which is expressed simultaneously to *feoB* 89% of the time (Lau et al. 2013), *feoC* is not conserved between species. *Trichodesmium* is an example where only homologues for *feoA* and *feoB* are encountered in the genome.

Mutants of both the Feo and Fbp/Fut uptake systems in cyanobacteria have reduced growth (Katoh et al. 2001a). Their expression, which is controlled by Fe concentrations, appears to be regulated by Fur in pathogenic bacteria (Mey et al. 2005; Ho & Ellermeier 2015) and *Synechocystis* (Krynická et al. 2014).

A common pathway of Fe uptake in phytoplankton involves reduction of Fe$^{3+}$ to Fe$^{2+}$ prior to its uptake by means of cell derived electrons (Roe & Barbeau 2014; Kranzler et al. 2014; Lamb et al. 2014) as a way of inducing Fe dissociation from ‘unpalatable’ ligands (Kranzler et al. 2011). Although some siderophores can be internalised through the Fec/Fhu ABC transport systems (Fig. 1.7), other substrates like dust particles are not available for direct uptake by the cells (Rich & Morel 1990; Rueter et al. 1992).

In pathogenic bacteria, yeast and eukaryotic phytoplankton reductive Fe uptake can occur through mechanisms like the export of reactive oxygen species (ROS) (Rose & Waite 2005) or the use of a ferric reductase (FRE) (Yun et al. 2001; Kustka et al. 2007; Groussman et al. 2015) (Fig. 1.7). In some cases reduction by FRE can be followed by reoxidation using multicopper oxidase (MCO) (Stearman et al. 1996; Askwith & Kaplan 1998; Herbik et al. 2002; Paz et al. 2007a; Groussman et al. 2015) to assist transport through associated specialised Fe$^{3+}$ transporters (Ftr1, TTf) (Stearman et al. 1996; Fisher et al. 1997, 1998; Urbanowski & Piper 1999; Paz et al. 2007b) (Fig. 1.7).

In cyanobacteria, including *Trichodesmium*, reactive oxygen species (ROS) mediated Fe reduction was previously shown (Rose & Waite 2005; Roe & Barbeau 2014; Hansel et al. 2016). Furthermore, *Synechocystis* is suggested to employ the plasma membrane located
alternative respiratory terminal oxidase (ARTO) for transfer of electrons to Fe³⁺ (Kranzler et al. 2014) (Fig. 1.7).

In addition to *Trichodesmium*’s cellular Fe uptake pathways, its colonial nature is potentially key to Fe acquisition *in situ*. Fe from desert dust was previously indicated to be more accessible to environmentally collected *Trichodesmium* puff colonies (Fig. 1.5Bi) compared to free filaments. Impressive footage by the same study recorded puff colony-associated filaments to engage in coordinated motility and shuffle dust from the periphery to the colony core (Rubin et al. 2011). Differences have been also identified between puff and tuft morphotypes of *Trichodesmium* in Fe uptake (Achilles et al. 2003) and ROS (superoxide) production (Hansel et al. 2016).

The ability of filamentous *Trichodesmium* in culture to acquire Fe from desert dust and the importance of cell surface contact for this process is further examined in Chapter 5 of this work. Also transcriptomic analysis targets the study of the organism’s distinct life styles when encountered with Fe deficient as opposed to Fe replete conditions.

1.4.3 Trophic life strategies- Genomic minimalism or maximalism?

There is a dichotomy between two distinct growth strategies, both concurrently employed in the ocean by different groups of microbes. Oligotrophy is characterised by stable slow growth designed to ensure persistence under nutrient deficient environments (Kirchman 2015). Maintenance of low biomass, even when encountered with increased nutrient abundance, is hypothesised to provide an advantage in avoidance of viral and grazing predation (a strategy referred to as ‘cryptic escape’) (Yooseph et al. 2010). Contrasting oligotrophs, copiotrophs have more flexible growth rates, that are adjusted depending on nutrient availability and can lead to blooms under improved trophic conditions (Yooseph et al. 2010).

This feast-or-famine lifestyle of the copiotrophs requires maintenance of larger genomes harbouring more genes involved in signal transduction, quorum sensing, motility and other processes associated with the organisms’ interaction with their environment (Yooseph et al. 2010). Also, numerous specific nutrient transport systems can be maintained as opposed to fewer more generalised transporters in oligotrophic organisms (Lauro et al. 2009). The genomic toolkit of copiotrophs however, potentially comes at a cost (either energetic or of other resources) compared to the streamlined genome of oligotrophs (Morris et al. 2012).
1.5 Omics and Detection of stress responses

Twenty years after Francis Crick first discussed the central dogma of biology, the unidirectional flow of information within an organism from DNA to RNA to protein (Crick 1958) (Fig. 1.7), sequencing techniques enabling decoding of nucleic acids became available (Maxam & Gilbert 1977; Sanger et al. 1977). They led to the sequencing revolution which started with the first complete genome sequence of *Haemophilus influenzae* (Fleischmann et al. 1995) and continued to currently reach more than 30,000 sequenced bacterial genomes (https://gold.jgi.doe.gov/, Reddy et al. 2015). Several techniques have since emerged to decode cellular transcriptomes, the DNA stored information that is expressed at a given moment in the form of messenger RNA (mRNA). Transcriptomic analysis was revolutionised with the development of the faster and cheaper next generation sequencing techniques and RNAseq in 2005 (Mardis 2011).

The snapshot of a cellular transcriptome in time is the result of a compilation of regulatory changes in transcription caused by environmental stimuli (Fig. 1.8). It involves the interaction of signalling pathways with DNA-binding transcription activators and repressors and sigma/antisigma factors regulating transcription by RNA polymerases (Wösten 1998). In addition, non-protein-coding RNA (ncRNA) modifications which can affect the nature of the transcript pool include mobile group-II introns (Ferat & Michel 1993) which interfere with transcription, and diversity generating retroelements (DGR) (Doulatov et al. 2004) introducing changes to the RNA nucleotide sequence. More recently, epigenetic regulation, where hereditary changes in the DNA physical structure, but not the nucleotide sequence, alter gene expression has also been explored (Blyn et al. 1990; Turner et al. 2009; Casadesus & Low 2013).

Transcription does not always equate to protein expression of the mRNA product since post-transcriptional controls also include RNA degradation, targeted through 5′-triphosphate removal and polyadenylation (Li et al. 2002; Kime et al. 2010) (Fig. 1.8). Although tools for the detection of the previously mentioned controls are available by RNAseq (Marguerat & Bähler 2010), modifications at the level of translation (Fig. 1.8) (Byrgazov et al. 2013; Sauert et al. 2015) are more challenging to identify through this technique and not very well characterised. The latter include controls such as ribosome, translation factor and tRNA heterogeneity which can modulate the rate of translation of different transcripts.
Chapter 1: Introduction

Figure 1.8: The cellular information flow and interactions with the environment. Environmental perturbations, natural or anthropogenic, stimulate gene expression changes. The transcribed RNA will form the cells transcriptome which will be subjected to post-transcriptional modification and subsequently translated to protein. Translational controls affect the rate and nature of translation, while the resulting proteome will be further modified after its generation. Protein activity will subsequently characterise the metabolome, which provides an indication of cellular activity (the phenome). Through this pathway processes, including photosynthesis and N₂ fixation, can be adjusted as a response to the environmental changes. Combined changes at the community level will affect the functioning of biogeochemical cycles and the consequently the environment (e.g. climate).
In recent years, the availability of sequenced genomes and the improvement of mass spectrometry through novel ionisation techniques facilitated proteomic studies which assemble information translated from mRNA to proteins (reviewed in Yates 2011). Further downstream the route of DNA to activity, information at the proteomic level is generally considered a better indicator of the cell’s physiological state. Although, the biggest control of protein expression still appears to be transcription itself (Li et al. 2014; Li & Biggin 2015; Jovanovic et al. 2015; Battle et al. 2015) this seems to be dependent on the stress state of the cells and the organism under investigation (Vogel & Marcotte 2012; Liu & Aebersold 2016; Cheng et al. 2016). For example, transcriptional control is suggested to be more important in organisms with copiotrophic rather than oligotrophic lifestyles (Cottrell & Kirchman 2016).

Post translational modifications including protein degradation mean that activity and protein abundance may not always correlate either, and therefore the field of metabolomics, quantitatively characterising cell produced metabolites has been developed to complement the omic analytic methods (Patti et al. 2012) (Fig. 1.8).

*Trichodesmium*’s genome was first published in 2003 (Joint Genome Institute, [http://genome.jgi.doe.gov/trier/trier.info.html](http://genome.jgi.doe.gov/trier/trier.info.html)). It is unusually large compared to other cyanobacteria genomes (Shih et al. 2013), being 7.75 Mbp in length and harbouring 4451 predicted genes, while it has a particularly high proportion (40%) of non-protein coding genome compared to the cyanobacterial average (15%) (Larsson et al. 2011). The recently published primary transcriptome of *Trichodesmium* (Pfreundt et al. 2014) revealed its large toolbox of regulatory DNA (Pfreundt et al. 2014; Pfreundt et al. 2015). Interestingly, *Trichodesmium*’s large non-coding genome content also appears to be maintained in the environment (Walworth et al. 2015) pointing out to an important function promoting its conservation in the natural conditions encountered by the organism *in situ*. The first omic study in *Trichodesmium* was performed in 2011 looking at the response of the proteome to different nitrogen regimes (Sandh et al. 2011) and was followed by the analysis of its proteomic response to Fe and P-limitation (Snow et al. 2015; Walworth et al. 2016).

Although omic techniques are currently revolutionising the field of environmental microbiology their potential for understanding the function of characterised genes/proteins is largely limited. To this effect, targeted approaches through ‘reverse genetics’ can provide functional characterisation by using gene engineering coupled with physiological studies. For this, gene deletion/overexpression can provide clues towards protein function by
stimulating phenotypic alterations. Such techniques have been previously used to study the P and Fe uptake pathways of cyanobacteria (Martínez et al. 2012; Kranzler et al. 2014). Furthermore, the use of reporter genes/proteins, such as green fluorescent protein (GFP) can aid characterisation of in vivo protein localisation and temporal patterns of expression (Schuergers et al. 2015; Sacharz et al. 2015). The tools are not limited to biological systems as the activity and substrate binding affinity of overexpressed/purified proteins can be also examined in vitro, and their crystal structure elucidated, to facilitate functional characterisation. For instance the Fe$^{3+}$ binding capacity of the *Synechocystis* FutA transporters has previously been demonstrated using this approach (Katoh 2001b; Badarau et al. 2008).

Currently, there are no established techniques to genetically modify *Trichodesmium*, therefore targeted approaches are largely limited. However, genes can be incorporated in the DNA of foreign organisms, in a process referred to as heterologous expression. Through recovery of the function of homologous deleted genes in the host organism, or generation of novel phenotypes, the targeted gene/genes can be characterised. Such a technique was previously utilised to indicate the functionality of *Trichodesmium*’s transcription regulator HetR in modulating heterocyst formation in *Anabaena* sp. strain PCC 7120 when the organism’s native homologue was deleted (Kim et al. 2011).

### 1.6 Significance of molecular data in ocean science

The expanding molecular understanding of marine microorganism metabolic processes provides the foundation for several previously unattainable applications.

For example, sensing via the use of biomarkers (Zehr et al. 2009) can identify cellular activity or stress at different oceanic locations (Saito et al. 2014), the presence/distribution of specialised populations of microorganisms such as diazotrophs (Robidart et al. 2014) and can provide an early warning system for detection of harmful species blooms (Tsaloglou et al. 2013).

In addition, a molecular understanding of cellular processes can be used to construct models of metabolism at the genome level (Genome scale metabolic network reconstructions) that are used to identify new metabolic capabilities within species (Levering et al. 2016). Gathered information about genes/proteins can be further
incorporated in models used to predict the biogeography, activity and response of organisms to environmental changes (Reed et al. 2014; Mock et al. 2016).

Finally, molecular tools can be employed to identify and utilise biotechnologically interesting proteins (Kennedy et al. 2011) such as biocatalysts (Parages et al. 2016), novel antibiotics (Eom et al. 2013) and biosurfactants (Gudiña et al. 2016) from the vast array of unstudied marine microorganism species.

### 1.7 Thesis plan

The work presented in this thesis employs a combination of techniques, both targeted and holistic, molecular and physiological to establish a better understanding of *Trichodesmium*’s nutrient limitation responses and uptake mechanisms. We focus on the two most discussed nutrients, P and Fe, both suggested to be limiting growth and activity of this important primary producer in the ocean (Moore et al. 2009; Mills et al. 2004).

In view of recent evidence pointing towards the abundance and significance of the reduced inorganic form of P, PO$_3^{3-}$ in the ocean (Pasek et al. 2014; Van Mooy et al. 2015), Chapter 3 is dedicated to elucidating *Trichodesmium*’s ability to utilise it for growth. The role of the organism’s *ptxABCD* homologues is established through their heterologous expression in *Synechocystis* and controls on their regulation are determined.

The same technique is employed in Chapter 4 to decipher the function of the FutA homologue in *Trichodesmium*. This protein, which has previously been used as a biomarker of Fe stress in the ocean (Rivers et al. 2009; Saito et al. 2014) is equally similar to two *Synechocystis* Fe regulated homologues performing distinct functions (Tölle et al. 2002; Waldron et al. 2007; Kranzler et al. 2014). Mutants of both *Synechocystis* homologues are acquired and the *Trichodesmium* gene is introduced in their genome to identify recovered phenotypes. Additionally, the cellular localisation of the *Trichodesmium* and *Synechocystis* homologues is explored through fusion of their signal sequences to green fluorescent protein (GFP). The *Trichodesmium* protein is further studied by overexpression and purification for structural studies aimed at assessing its metal binding capacity.

Finally, in Chapter 5 the ability of *Trichodesmium* to access Fe provided as Saharan desert dust and the importance of cell surface contact for this process is established through specifically designed physiological experiments. The organism’s transcriptomic response
when grown under the variable Fe regimes used during the physiological experiments is analysed to detect the effect on known *Trichodesmium* Fe metabolism coordinating genes, and identify novel ones.
Chapter 2
Materials and methods

2.1 Strains and growth conditions

Strains of *Synechocystis* sp. PCC 6803 (*Synechocystis*), *Trichodesmium erythraeum* IMS101 (*Trichodesmium*) and *Escherichia coli* (*E.coli*) generated and used throughout this work are detailed in Table 2A.

2.1.1 *E. coli*

*Escherichia coli* was grown at 37°C in LB (Luria-Bertani) broth in 50 ml glass conical flasks or 15 ml falcon tubes (Corning Inc., NY, USA), and plates (Fisher Scientific, Loughborough, UK) set with 1% (w/v) agar, with or without 100 µg/ml ampicillin.

2.1.2 *Trichodesmium*

*Trichodesmium* was grown at 27 °C in modified YBC-II media (Table 2B, 2C) (Chen 1996), (referred to as mYBC-II). The molybdenum content of the media was raised to 100 nM to better match seawater concentrations (James 2005) while other metals were modified based on trace metal additions by Shi et al. (2012). The media was filter sterilised using 0.22 µm Stericup Filter units with 1 L Sterile Receiver flasks (Millipore, MA, USA). Cells were grown in 25 cm² sterile polystyrene cell culture flasks with 0.2 µm vent caps (Corning Inc., NY, USA) subjected to gentle orbital shaking (150 rpm) under a 12 hr/12 hr light (ca. 130 µmol photons m⁻² s⁻¹)/dark cycle.
Table 2A: Bacterial strains used for experiments. Gene insertions/deletions in *Synechocystis* were selected using antibiotic resistance to kanamycin (kan), chloramphenicol (cm) and zeocin (zn). Plasmid selection in *E. coli* was conferred using ampicillin (amp) resistance.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Gene deletion (Ant&lt;sup&gt;β&lt;/sup&gt; µg/ml)</th>
<th>Gene insertion (Ant&lt;sup&gt;β&lt;/sup&gt; µg/ml)</th>
<th>Source</th>
<th>Chapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT-G</td>
<td></td>
<td>Tery_0365-0368 (kan30)</td>
<td>Cereda et al. 2014</td>
<td>3,4</td>
</tr>
<tr>
<td>PbxA8CD&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
<td>Tery_0365-0367 (kan30)</td>
<td>This study</td>
<td>3</td>
</tr>
<tr>
<td>PbxA3C&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
<td>Tery_0365-0367 (kan30)</td>
<td>This study</td>
<td>3</td>
</tr>
<tr>
<td>PbxD&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
<td>Tery_0368 (kan30)</td>
<td>This study</td>
<td>3</td>
</tr>
<tr>
<td>ΔfutA1</td>
<td>slr1295 (zeo20)</td>
<td></td>
<td>This study</td>
<td>4</td>
</tr>
<tr>
<td>ΔfutA2</td>
<td>slr0513 (cm25)</td>
<td></td>
<td>This study</td>
<td>4</td>
</tr>
<tr>
<td>ΔfutA1/ΔfutA2</td>
<td>slr1295/slr0513 (zeo20/cm25)</td>
<td></td>
<td>This study</td>
<td>4</td>
</tr>
<tr>
<td>ΔfutA1&lt;sup&gt;+&lt;/sup&gt;3377</td>
<td>slr1295 (eco20)</td>
<td>Tery_3377 (kan30)</td>
<td>This study</td>
<td>4</td>
</tr>
<tr>
<td>ΔfutA2&lt;sup&gt;+&lt;/sup&gt;3377</td>
<td>slr0513 (cm25)</td>
<td>Tery_3377 (kan30)</td>
<td>This study</td>
<td>4</td>
</tr>
<tr>
<td>FutA1ss-GFP</td>
<td>slr1295ss-GFP (cm25)</td>
<td></td>
<td>This study</td>
<td>4</td>
</tr>
<tr>
<td>FutA2ss-GFP</td>
<td>slr0513ss-GFP (cm25)</td>
<td></td>
<td>This study</td>
<td>4</td>
</tr>
<tr>
<td>3377ss-GFP</td>
<td>Tery_3377ss-GFP (cm25)</td>
<td></td>
<td>This study</td>
<td>4</td>
</tr>
<tr>
<td>3377ss-KK-GFP</td>
<td>Tery_3377ss (R5/6K) GFP (cm25)</td>
<td></td>
<td>This study</td>
<td>4</td>
</tr>
<tr>
<td><em>Trichodesmium erythraeum</em> IMS101</td>
<td></td>
<td></td>
<td>CCMP1985 (NCMA, Bigelow Laboratory for Ocean Sciences)</td>
<td>3,4,5</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td></td>
<td></td>
<td></td>
<td>3,4</td>
</tr>
<tr>
<td>XL1-Blue</td>
<td></td>
<td></td>
<td>Agilent Technologies, Stockport, UK</td>
<td>5,4</td>
</tr>
<tr>
<td>BL21(&lt;DEL3)</td>
<td></td>
<td></td>
<td>New England Biolabs Ltd, Hitchin, UK</td>
<td>4</td>
</tr>
<tr>
<td>DH5-alpha</td>
<td></td>
<td></td>
<td>New England Biolabs Ltd, Hitchin, UK</td>
<td>4</td>
</tr>
<tr>
<td>3377 OE</td>
<td></td>
<td>Tery_3377 (in pET21a+) (amp100)</td>
<td>This study</td>
<td>4</td>
</tr>
</tbody>
</table>
### Table 2B: Modified (m) YBC-II media recipe

<table>
<thead>
<tr>
<th>YBC-II medium</th>
<th>Per L H$_2$O</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vitamins</strong></td>
<td></td>
</tr>
<tr>
<td>1000x Vitamin stock</td>
<td>1 ml</td>
</tr>
<tr>
<td><strong>Trace metals</strong></td>
<td></td>
</tr>
<tr>
<td>1000x Iron stock</td>
<td>1 ml (Standard growth)</td>
</tr>
<tr>
<td></td>
<td>2.5 ml (Fe+ experimental cultures)</td>
</tr>
<tr>
<td><strong>Nutrients</strong></td>
<td></td>
</tr>
<tr>
<td>1000x Phosphorus stock:</td>
<td></td>
</tr>
<tr>
<td>Phosphate stock</td>
<td>1 ml (Standard growth)</td>
</tr>
<tr>
<td>Phosphite stock</td>
<td>1 ml (Chapter 3- PO$_3^3$ cultures)</td>
</tr>
<tr>
<td><strong>Salts</strong></td>
<td></td>
</tr>
<tr>
<td>Sodium chloride</td>
<td>24.55 g</td>
</tr>
<tr>
<td>Potassium chloride</td>
<td>0.75 g</td>
</tr>
<tr>
<td>Sodium bicarbonate</td>
<td>0.21 g</td>
</tr>
<tr>
<td>Boric acid</td>
<td>0.036 g</td>
</tr>
<tr>
<td>Potassium bromide</td>
<td>0.116 g</td>
</tr>
<tr>
<td>Magnesium chloride</td>
<td>4.067 g</td>
</tr>
<tr>
<td>Calcium chloride</td>
<td>1.47 g</td>
</tr>
<tr>
<td>Magnesium sulphate</td>
<td>6.16 g</td>
</tr>
<tr>
<td>Sodium fluoride stock</td>
<td>1 ml</td>
</tr>
<tr>
<td>Strontium chloride stock</td>
<td>1 ml</td>
</tr>
<tr>
<td><strong>Stock solutions</strong></td>
<td>Per L H$_2$O</td>
</tr>
<tr>
<td><strong>Salts</strong></td>
<td></td>
</tr>
<tr>
<td>1000x Sodium fluoride stock</td>
<td>2.939 g</td>
</tr>
<tr>
<td>1000x Strontium chloride stock</td>
<td>0.017 g</td>
</tr>
<tr>
<td><strong>Trace metals</strong></td>
<td></td>
</tr>
<tr>
<td>1000x Trace metal stock:</td>
<td></td>
</tr>
<tr>
<td>Manganese chloride</td>
<td>0.00396 g</td>
</tr>
<tr>
<td>Zinc Sulphate</td>
<td>0.00115 g</td>
</tr>
<tr>
<td>Cobalt (II) chloride</td>
<td>0.00059 g</td>
</tr>
<tr>
<td>Sodium molybdate</td>
<td>0.00266 g</td>
</tr>
<tr>
<td>Copper (II) sulphate</td>
<td>0.00025 g</td>
</tr>
<tr>
<td>Disodium EDTA</td>
<td>0.74448 g</td>
</tr>
<tr>
<td>1000x Iron stock</td>
<td>0.0432 g</td>
</tr>
<tr>
<td>Iron (III) chloride</td>
<td>0.0596 g</td>
</tr>
<tr>
<td>Disodium EDTA</td>
<td></td>
</tr>
<tr>
<td><strong>Nutrients</strong></td>
<td></td>
</tr>
<tr>
<td>1000x Phosphate stock</td>
<td>6.899 g</td>
</tr>
<tr>
<td>Monosodium phosphate</td>
<td></td>
</tr>
<tr>
<td>1000x Phosphite stock</td>
<td>10.802 g</td>
</tr>
<tr>
<td>Sodium phosphite</td>
<td></td>
</tr>
<tr>
<td><strong>Vitamins</strong></td>
<td></td>
</tr>
<tr>
<td>1000x Vitamin stock:</td>
<td></td>
</tr>
<tr>
<td>Thiamine</td>
<td>0.0998 g</td>
</tr>
<tr>
<td>Biotin</td>
<td>0.0005 g</td>
</tr>
<tr>
<td>Cyanocobalamin</td>
<td>0.0005 g</td>
</tr>
</tbody>
</table>
Table 2C: Modified YBC-II trace metal composition compared to standard YBC-II.

<table>
<thead>
<tr>
<th></th>
<th>Modified YBC-II</th>
<th>Standard YBC-II</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDTA</td>
<td>20 μM</td>
<td>2 μM</td>
</tr>
<tr>
<td>Cu</td>
<td>8 nM</td>
<td>1 nM</td>
</tr>
<tr>
<td>Zn</td>
<td>20 nM</td>
<td>4 nM</td>
</tr>
<tr>
<td>Co</td>
<td>8 nM</td>
<td>2.5 nM</td>
</tr>
<tr>
<td>Mn</td>
<td>18 nM</td>
<td>20 nM</td>
</tr>
<tr>
<td>Mo</td>
<td>100 nM</td>
<td>11 nM</td>
</tr>
<tr>
<td>Ni</td>
<td>20 nM</td>
<td>-</td>
</tr>
<tr>
<td>Se</td>
<td>10 nM</td>
<td>-</td>
</tr>
</tbody>
</table>

Cultures were routinely grown with 160 nM iron (FeCl$_3$-EDTA). For iron (Fe)-stress growth experiments (Chapter 5) no Fe was added to Fe-deplete treatments (Fe-) compared to 400 nM of added Fe to Fe-replete treatments (Fe+).

Saharan desert dust was collected between 04°43’S, 28°55’W and 06°56’N, 28°07’ using a mesh system aboard the RSS Shackleton on 01/04/1981-02/04/1981 (Murphy 1985). Equal amounts (0.25 mg ml$^{-1}$) were added to the media either directly or in 8 kDa MWCO dialysis tubing (DT) (BioDesign, NY, USA). Immediately before use, DT was tied on one end and boiled twice in 120 mM Na$_2$HCO$_3$ followed by another two times in 10 mM Na$_2$EDTA and 10 mM NaOH. After that, DT was rinsed three times in milli-Q H$_2$O (method adjusted from http://alta.biology.utah.edu/index.php, Horvath Lab Protocols Department of Biology, University of Utah). Clean DT was then handled with 10% HCl washed plastic tweezers and was included in all experimental treatments except form a no-DT control.

For *Trichodesmium* phosphorus (P)-stress growth experiments (Chapter 3) 50 μM phosphate (NaH$_2$PO$_4$.H$_2$O) (PO$_4^{3-}$) or phosphite (Na$_2$HPO$_3$.5H$_2$O) (PO$_3^{3-}$) were used as P sources while no P was added to control treatments.

Treatments for both, Fe (Chapter 5) and P (Chapter 3), experiments were set in triplicate cultures and inoculations were performed using cells grown to late exponential phase and pre-concentrated by filtering onto Nucleopore™ track-etched polycarbonate membranes (Whatman, Kent, UK) with a pore size of 5μm, washed with and resuspended in Fe/P free mYBC-II.
2.1.3 Synechocystis

Synechocystis was routinely grown in BG-11 media (Table 2D) (Rippka et al. 1979) at 30°C, shaking at 150 rpm, under constant illumination of ca. 50 μmol photons m² s⁻¹. BG-11 was supplemented with 5 mM glucose and buffered to pH 8.2 with 10 mM N-17 [tris(hydroxymethyl)methyl]-2-aminoethanesulfonic acid (TES)-KOH buffer. For growth on plates, 1% agar was added. Antibiotics were used in the concentrations indicated (Table 2A).

Table 2D: BG-11 media recipe

<table>
<thead>
<tr>
<th>BG-11 medium</th>
<th>Per L H₂O</th>
</tr>
</thead>
<tbody>
<tr>
<td>100x BG-11 stock</td>
<td>10 ml</td>
</tr>
<tr>
<td>1000x Iron stock</td>
<td>1 ml</td>
</tr>
<tr>
<td>1000x Phosphate stock</td>
<td>1 ml</td>
</tr>
<tr>
<td>1000x Carbonate stock</td>
<td>1 ml</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Supplementary stocks</th>
<th>Per L H₂O</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose stock</td>
<td>5 ml</td>
</tr>
<tr>
<td>TES/KOH</td>
<td>100 ml</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stock solutions</th>
<th>Per L H₂O</th>
</tr>
</thead>
<tbody>
<tr>
<td>100x BG-11 stock</td>
<td></td>
</tr>
<tr>
<td>Sodium nitrate</td>
<td>149.6 g</td>
</tr>
<tr>
<td>Magnesium sulphate</td>
<td>7.49 g</td>
</tr>
<tr>
<td>Calcium chloride</td>
<td>3.60 g</td>
</tr>
<tr>
<td>Citric acid</td>
<td>0.60 g</td>
</tr>
<tr>
<td>Disodium EDTA</td>
<td>0.56 ml of 0.5M stock pH 8.0</td>
</tr>
<tr>
<td>Trace mineral stock</td>
<td>100 ml</td>
</tr>
</tbody>
</table>

**1000x Iron stock**
- Ferric ammonium citrate 6 g

**1000x Phosphate stock**
- Dipotassium hydrophosphate 30.5 g

**1000x Phosphite stock**
- Sodium phosphate 37.81 g

**1000x Carbonate stock**
- Sodium carbonate 20 g

**1000x Iron stock**
- Trace mineral stock
  - Boric acid 2.86 g
  - Magnesium sulphate 1.81 g
  - Zinc sulphate 0.22 g
  - Sodium molybdate 0.39 g
  - Copper sulphate 0.079 g
  - Cobaltous nitrate 0.049 g

**1M Glucose stock**
- Glucose 180 g

**1M TES/KOH**
- TES 229.2 g, pH 8.2
Iron-stress growth experiments (Chapter 4) were carried out in phototrophic conditions (no glucose), in triplicate cultures, using YBG-11 media (recipe described by Shcolnick et al. 2007) with 6 mM or 0.6 mM of FeCl\textsubscript{3} for Fe-replete (Fe+) and -deplete (Fe-) conditions respectively.

Ferrozine (FZ) growth experiments were carried out in Fe- (0.6 mM of FeCl\textsubscript{3}) YBG-11 at a FZ final concentration of 200 μM (FZ+) or no added FZ (FZ-) (Chapter 4). Cultures were incubated at ca. 30 μmol photons m\textsuperscript{-2} s\textsuperscript{-1} of red light to minimise photoreduction of Fe (Kranzler et al. 2011).

For phosphorus (P) experiments (Chapter 3) triplicate cultures with 175 μM sodium phosphate (Na\textsubscript{2}H\textsubscript{3}PO\textsubscript{4}.H\textsubscript{2}O) (PO\textsubscript{4}\textsuperscript{3-}) or the same concentration of sodium phosphite (Na\textsubscript{2}HPO\textsubscript{3}.5H\textsubscript{2}O) (PO\textsubscript{3}\textsuperscript{3-}) were compared to no P added controls.

### 2.2 Physiological measurements

#### 2.2.1 Photosynthetic physiology

Photosynthetic physiology was monitored using a FASTtracka MkII Fast Repetition Rate fluorometer (FRRf) integrated with a FastActTM Laboratory system (Chelsea Technologies Group Ltd, Surrey, UK). For *Trichodesmium* measurements were made 2.5 hrs after the beginning of the photoperiod. Both *Trichodesmium* and *Synechocystis* samples were dark adapted for 20 min and consequently exposed to a background irradiance of 29 μmol photons m\textsuperscript{-2} s\textsuperscript{-1} (PAR) (Table 2E) for 2-5 min prior to measurements. This stimulates an increase in quantum yield (Richier et al. 2012).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequences per acquisition</td>
<td>12</td>
</tr>
<tr>
<td>Sequence interval</td>
<td>100</td>
</tr>
<tr>
<td>Acquisition pitch (s)</td>
<td>2</td>
</tr>
<tr>
<td>Flashlets in sat. phase</td>
<td>200</td>
</tr>
<tr>
<td>Flashlet duration in sat. phase (ms)</td>
<td>1</td>
</tr>
<tr>
<td>Flashlets in rel. phase</td>
<td>30</td>
</tr>
<tr>
<td>Flashlet duration in rel. phase (ms)</td>
<td>49</td>
</tr>
<tr>
<td>Saturation phase duration (ms)</td>
<td>0.4</td>
</tr>
<tr>
<td>Relaxation phase duration (ms)</td>
<td>57.7</td>
</tr>
<tr>
<td>LED set</td>
<td>160</td>
</tr>
<tr>
<td>Background irradiance (μmol photons m\textsuperscript{-2} s\textsuperscript{-1})</td>
<td>29</td>
</tr>
</tbody>
</table>
$F_v/F_m$ was used as an estimate of the apparent PSII photochemical quantum efficiency (Kolber et al. 1998). The dark adapted state ($F_o$) and maximum ($F_m$) fluorescence were blank corrected, and the difference between them ($F_v$) was recalculated to manually compute an adjusted $F_v/F_m$ ratio. The functional absorption cross-section of PSII ($\sigma_{PSII}$) was also acquired (Table 2F). Data presented is the average of three technical replicates or a single measurement (for *Trichodesmium* and *Synechocystis* respectively) for each of three independent cultures (biological replicates).

<table>
<thead>
<tr>
<th>Technique</th>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photosynthetic physiology</td>
<td>$F_m$</td>
<td>The maximum fluorescence at saturation. All PSII reaction centres are ‘closed’.</td>
</tr>
<tr>
<td></td>
<td>$F_o$</td>
<td>Fluorescence at time=0. All PSII reaction centres are ‘open’.</td>
</tr>
<tr>
<td></td>
<td>$F_v$</td>
<td>The difference between the maximum fluorescence ($F_m$) and initial fluorescence ($F_o$).</td>
</tr>
<tr>
<td></td>
<td>$F_v/F_m$</td>
<td>Apparent PSII photochemical quantum efficiency</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{PSII}$ (nm$^2$)</td>
<td>Functional absorption cross-section of PSII</td>
</tr>
</tbody>
</table>

**Table 2F**: Description of physiological parameters recorded.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth (Trichodesmium)</td>
<td>Cell number</td>
<td>$C_N = \sum f_l$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$C_N =$ cell number</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\Sigma f_l =$ total filament length</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$C_l =$ cell length</td>
</tr>
</tbody>
</table>

Growth (Synechocystis) | OD | Measured at a wavelength of 750 nm where no pigment absorption interference occurs. |
| | | An OD$_{750} = 1$ is approximately equal to 1.6$\times$10$^8$ cells ml$^{-1}$ (Pojidaeva et al. 2004) |

| Growth rate | $\mu = \frac{\ln(cC_{t_2}/cC_{t_1})}{t_2 - t_1}$ |
| | $cC_t =$ cell concentration (direct measurement or using OD) |
| | $t =$ time |

2.2.2 Growth

*Trichodesmium* growth was monitored through cell counts performed using a Sedgewick rafter counting chamber and a GX CAM-1.3 camera on an L1000A biological microscope (GT Vision Ltd, Suffolk, UK). Cell and filament lengths were identified using GX capture software (GX 14 Optical, Suffolk, UK) and ImageJ software (Schneider et al. 2012). The number of cells per ml was calculated by dividing the total filament length by the average cell length (Table 2F).
Chapter 2: Materials and Methods

*Synechocystis* growth was monitored turbidimetrically by recording the optical density of the cultures at a wavelength of 750 nm (OD\textsubscript{750}) using a Jenway 7315 spectrophotometer (Bibby Scientific Ltd, Staffordshire, UK). When OD\textsubscript{750} exceeded that of 1, samples were diluted with filtered medium to achieve better measurement accuracy.

For both *Trichodesmium* and *Synechocystis*, growth rates were calculated from the gradient of the natural logarithm of measurements during exponential growth (Table 2F).

### 2.2.3 Chlorophyll \(\alpha\)

Chlorophyll \(\alpha\) (Chl \(\alpha\)) content was assessed through extractions from 2 ml *Trichodesmium* cell culture filtered on MF 300 microfibre filters (Fisher Scientific, Loughborough, UK) or 100-200 \(\mu\)l *Synechocystis* culture. Cells were incubated in 90% acetone for 24-48 hrs in the dark, at 4 °C. Fluorometric measurements were recorded on a Turner TD-700 fluorometer (Turner Designs, CA, USA) as described by Welschmeyer (1994). The results were calibrated using solid secondary standards and Chl \(\alpha\) concentration was calculated using a Chl \(\alpha\) standard curve.

### 2.2.4 Statistical analysis

Statistical differences in physiological parameters were assessed using the Student’s t-test (Chapter 3) or general linear model (GLM) followed by a post-hoc Tukey-test (CI=95%) (Chapters 4 and 5) (Minitab 17.3.1, Minitab Inc., Coventry, UK).

### 2.3 Methods for molecular analysis

#### 2.3.1 Nucleic acid extractions and quality control

**RNA extractions:** RNA was extracted using the spin column based RNeasy Plant Mini Kit (Qiagen, Manchester, UK) according to the manufacturers’ instructions subjected to minor alterations as follows: Cells (~ 5 \(\times\) 10\(^6\)) were filtered on MF 300 microfibre filters (*Trichodesmium*) or pelleted by centrifugation (*Synechocystis*) and snap frozen in liquid nitrogen. They were subsequently stored at -80 °C for future use or processed for RNA extractions immediately. Homogenization was achieved by disruption with a pestle and resuspension in RLT lysis buffer. After removal of cell and filter-paper debris by centrifugation for 2 mins through the QIAshredder spin column, samples were loaded on the RNeasy spin column and 96-100% ethanol was used to wash the spin column twice or
until the flow through was colourless. The manufactures’ protocol was followed thereafter and RNA was eluded in 30 μl nuclease free water. Extracted RNA was stored at -80 °C or DNase treated directly (Sec. 2.3.2).

**DNA extractions:** DNA from both *Trichodesmium* and *Synechocystis* was isolated using the DNeasy Plant Mini Kit (Qiagen, Manchester, UK) according to the manufacturers’ instructions. Homogenisation was achieved as described above and the manufacturer’s guidelines were followed for the subsequent extraction. Purified DNA was stored at -20 °C in buffer AE.

**Nucleic acid quality control:** Purity and initial quantification of 1 μl extracted DNA and RNA was determined from the 260/280 and 260/230 nm ratios measured on a NanoDrop ND-1000 UV-Vis Spectrophotometer (Thermo Scientific, DE, USA). The nucleic acid 260/280 value of ~1.8 for DNA or ~2 for RNA is generally considered a sign of purity implied by the absence of protein, phenol or other 280 nm absorbing contaminants. Equally as a secondary measure of purity a 260/230 ratio between 2.0 and 2.2 is expected. However, contaminants that absorb at 230 nm (lowering the ratio) can include guanidine thiocyanate a lysis buffer ingredient which at subnanomolar concentrations can decrease the 260/230 considerably without compromising the performance of the quantitative PCR reactions (VonAhlfen & Schlumpberger 2010). The absence of potential inhibition from any contaminants was also confirmed through linearity of quantitative Polymerase Chain Reaction (qPCR) standard curves (Sec. 2.3.3).

Integrity of RNA was determined from the electropherogram trace, RNA integrity number (RIN) and percentage of RNA fragments above 200 nucleotides (DV200) as produced by the Agilent 2100 Bioanalyzer system using the Agilent RNA 6000 Nano Kit (Agilent Technologies UK Limited, Cheshire, UK) (Table 2G).
Table 2G: RNA extracted-quality control and concentrations.

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Treatment</th>
<th>Sample</th>
<th>260/280</th>
<th>260/230</th>
<th>RIN</th>
<th>DV200 (%)</th>
<th>Concentration (ng/μl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>P-</td>
<td>1</td>
<td>1.97</td>
<td>1.35</td>
<td>6.6</td>
<td>95</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>1.99</td>
<td>1.36</td>
<td>6.8</td>
<td>93</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>2.00</td>
<td>1.54</td>
<td>6.6</td>
<td>94</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>PO₄⁻</td>
<td>1</td>
<td>2.06</td>
<td>1.37</td>
<td>6.7</td>
<td>93</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>2.00</td>
<td>0.96</td>
<td>6.7</td>
<td>90</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>1.97</td>
<td>1.58</td>
<td>6.6</td>
<td>90</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>PO₄⁺</td>
<td>1</td>
<td>2.04</td>
<td>1.47</td>
<td>5.7</td>
<td>91</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>1.98</td>
<td>1.64</td>
<td>6.1</td>
<td>97</td>
<td>191</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>2.03</td>
<td>1.43</td>
<td>8</td>
<td>97</td>
<td>72</td>
</tr>
<tr>
<td>4</td>
<td>Fe⁻</td>
<td>1</td>
<td>2.06</td>
<td>2.24</td>
<td>7.4</td>
<td>97</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>1.99</td>
<td>1.91</td>
<td>7.8</td>
<td>99</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>1.97</td>
<td>1.87</td>
<td>8</td>
<td>95</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Fe⁺</td>
<td>1</td>
<td>2.07</td>
<td>2.30</td>
<td>7.6</td>
<td>98</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>2.07</td>
<td>2.35</td>
<td>7.6</td>
<td>99</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>2.03</td>
<td>2.09</td>
<td>7.4</td>
<td>99</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>[Dust]</td>
<td>1</td>
<td>2.06</td>
<td>2.17</td>
<td>7.6</td>
<td>99</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>1.93</td>
<td>1.28</td>
<td>7.4</td>
<td>98</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>2.00</td>
<td>1.57</td>
<td>7.5</td>
<td>97</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Dust⁺</td>
<td>1</td>
<td>2.04</td>
<td>2.09</td>
<td>7.4</td>
<td>97</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>2.02</td>
<td>2.3</td>
<td>7.5</td>
<td>98</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>2.06</td>
<td>2.32</td>
<td>7.5</td>
<td>99</td>
<td>37</td>
</tr>
</tbody>
</table>

2.3.2 DNase Treatment and cDNA synthesis

Since on-column DNase digestion or RNA clean-up using the RNeasy Plant Mini Kit was not effective in eliminating all identifiable traces of genomic DNA (gDNA) contaminants in trial RNA extractions, the isolated RNA was DNase treated with TURBO DNA-free DNase (Life Technologies Ltd, Paisley, UK) instead. After treatment RNA was stored at -80 °C or immediately converted to complementary DNA (cDNA) using random hexamers provided by the Tetro cDNA Synthesis Kit (Bioline Reagents Limited, London, UK). Complementary DNA (cDNA) was aliquoted to single use volumes and stored at -20 °C. Samples were tested for the absence of gDNA contamination by performing the synthesis reaction with and without Reverse Transcriptase (RT) for each sample. Absence of visible products after 40 cycles of PCR amplification in no RT controls using reference gene primers (see below) was indicative of below the detection limit or no gDNA contamination.
2.3.3 Polymerase Chain Reaction (PCR) conditions

**Primer design:** Oligonucleotide primers were designed using Primer3 through the Primer-BLAST tool on the NCBI portal ([http://www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) (Ye et al. 2012) and supplied from biomers.net GmbH (Ulm, Germany) or Integrated DNA Technologies Inc. (IA, USA). Lyophilised primers were rehydrated in 10 mM Tris-Cl, 0.1 mM EDTA, pH 8.5 to a concentration of 100 μM and stored at -20 °C.

For qPCR, 20-24 bp primers for amplification of 112-189 bp products were designed. Primer melting temperatures were chosen to be 60 ± 1 °C, the guanine (G) – cytosine (C) content between 30-80% and their last five bases at the 3’ end to contain the lowest number of G/C bases possible. Primers were diluted to 250 nM final concentration immediately before use. To ensure compliance with the MIQE guidelines (Bustin et al. 2009) a dilution series (≥4 data points) of cDNA sample was constructed and all primer sets were run in triplicate at each cDNA concentration. A qPCR reaction amplification efficiency between 90 and 105% and linearity greater than $r^2=0.97$ was termed satisfactory for further analysis (Pfaffl et al. 2004; Bustin et al. 2009; Bustin 2010) (Table 2H). At the end of each run, melt (dissociation) curve analysis was used to confirm single product amplification by primer sets.

**End point analysis:** End point reverse transcription (RT)-PCR was performed on a SureCycler 8800 Thermal Cycler (Agilent Technologies UK Limited, Cheshire, UK) and MyCycler Thermal Cycler (Bio-Rad Laboratories Limited, Hertfordshire, UK) with MyTaq High Specificity Red DNA Polymerase Mix (Bioline, London, UK) or Q5 High-Fidelity Polymerase Mix (New England Biolabs Ltd, Hitchin, UK) according to the manufactures’ instructions. Products were visualised using agarose gel electrophoresis (Sec. 2.3.4).

**Quantitative (q) PCR:** For Chapter 3 analysis was performed on an Mx3005P qPCR using Brilliant III Ultra-Fast SYBR Green qPCR Master mix (Agilent Technologies UK Limited, Cheshire, UK) at a total reaction volume of 20 μl. The PCR conditions were as follows: 1 cycle 95 °C for 180 secs (initial denaturation), 40 cycles of 95 °C for 15 secs (denaturation) followed by 60 °C for 20 secs (annealing/extension) and 1 cycle of 95 °C for 60 secs (final extension). Dissociation curves were created starting at 55 °C for 30 secs and increasing to 95°C.

For validation of transcriptomic results (Chapter 5) a ROCHE LightCycler 96 was used. Reactions were set at 20 μl with iQ SYBR Green qPCR Supermix (Bio-Rad Laboratories Limited, Hertfordshire, UK): 1 cycle 95 °C for 180 secs (initial denaturation), 45 cycles 95 °C.
for 10 secs (denaturation) followed by 60 °C for 30 secs (annealing/extension), 1 cycle of 95 °C for 10 secs (final extension). Dissociation curves were recorded between 55 °C and 95 °C.

For all primer sets each sample was run in triplicate (technical replicates) along with a triplicate set of no-template controls (NTC). This contained all reaction constituents apart from the sample in the place of which nuclease free water was added. For each gene analysed all samples of all treatments were included on a single plate therefore no inter-run calibrators (IRC) were required as a control of potential intra-run variation.

For Chapter 3 reference genes *rnpB* (Chappell & Webb 2010) and *rotA* (Orchard et al. 2009) were used. For Chapter 5 genes Tery_R0021 (*rnpB*), Tery_2660 (*GlyA*) (Chappell & Webb 2010), Tery_2329 and Tery_3279 (identified from transcriptomic analysis performed in the context of this work) were tested on geNorm (qBase+ software, Biogazelle, Zwijnaarde, Belgium). The best normalisation strategy predicted by the software suggested the use of Tery_2329 and Tery_3279 as reference genes. These produced a gene expression stability value (M) of 0.143 and a coefficient of variation on the normalized reference gene relative quantities (CV) of 0.05 well below the qBASE default thresholds of 0.5 and 0.2 respectively.

Normalised relative quantities (NRQs) in Chapter 3 were scaled using the phosphate-added (PO₄³⁻) treatment as control, while in Chapter 5 the lowest NRQ for the specific gene analysed. They were subsequently presented as relative fold change (RFC) (qBase+ software, Biogazelle, Zwijnaarde, Belgium). The built-in statistical wizard was used to compare the means of different treatments using one-way analysis of variance (ANOVA).

### 2.3.4 Agarose gel electrophoresis

All PCR and cloning products were visualised through agarose gel electrophoresis. Gels were made with 30 ml TAE (tris-acetate-EDTA) buffer, 3 μl (of 10 mg/ml stock) ethidium bromide and set with 0.3 g, 0.45 g or 0.8 g of agarose for 1%, 1.5% and 3% gels (the starting concentration for the latter is lower than 3% to allow for evaporation of liquid during the heating process) respectively. Gels were run in TAE buffer (supplemented with 3 μl of 10 mg/ml ethidium bromide) for 30-45 mins at 75-90 V. Molecular weight markers HyperLadder I, II or V (Bioline, London, UK) were used for 1, 1.5 and 3% gels respectively.
### Table 2H: Genes analysed by qPCR, amplicon and primer information

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Gene</th>
<th>Tery no.</th>
<th>Forward</th>
<th>Reverse</th>
<th>Size</th>
<th>Efficiency</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>ptxB</td>
<td>0366</td>
<td>GCGATCAAGCTTACTTTAGTC</td>
<td>CATTTCGCAACAGCATAAGCAAC</td>
<td>132</td>
<td>93.1</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>ptxD</td>
<td>0368</td>
<td>CCGACTTATTAGGAGCACCC</td>
<td>AGCTTTTCTAAATTCTCTCCATCC</td>
<td>189</td>
<td>99.3</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>rotA</td>
<td>0301</td>
<td>AGGAGGTTGGCCAAAGGTT</td>
<td>GCCACCTGTATCTTTACCAGCA</td>
<td>125</td>
<td>93</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>rnpB</td>
<td>R0021</td>
<td>GGAACCAGAAAGAAGACCAACC</td>
<td>GCCACAGAAAGAATACCGCACC</td>
<td>168</td>
<td>99.4</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>isiA</td>
<td>1667</td>
<td>TTGCTTGGCCTTGACCTGGG</td>
<td>AGGGAGGCAAGGCTGAGAAAC</td>
<td>145</td>
<td>92</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>nif-like</td>
<td>4114</td>
<td>GCTGTAATGGTGCTTGTTGC</td>
<td>TCCGCTTCCTCTCTTAGCTGT</td>
<td>143</td>
<td>93</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>cytc6</td>
<td>2561</td>
<td>TCTCCGCTGCAACATACCA</td>
<td>AGTGGAAGGGGGTAGATCCA</td>
<td>124</td>
<td>97</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>glyA</td>
<td>2660</td>
<td>GGCAGTGCTCCTCCTACCTAG</td>
<td>GCTTCCCAAAATGGCACTGC</td>
<td>125</td>
<td>96</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>gvpN</td>
<td>3292</td>
<td>ACCAGGTTGGTTGACAATTGGA</td>
<td>CACTTTGCGCAACACACTGC</td>
<td>134</td>
<td>96</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>RNase T2</td>
<td>3279</td>
<td>AGCTGGCCAACCGATTCTTG</td>
<td>TGTCAGCGCAAAACATGTA</td>
<td>112</td>
<td>97</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>rnpB</td>
<td>R0021</td>
<td>GGAACCAGAAAGAAGACCAACC</td>
<td>GCCACAGAAAGAATACCGCACC</td>
<td>168</td>
<td>100.1</td>
<td>1</td>
</tr>
</tbody>
</table>
2.4 Generation of genetically modified strains

2.4.1 Approaches

For functional characterisation of proteins, *Synechocystis* was genetically engineered to generate strains with genes of interest deleted (Chapter 4) and to heterologously express *Trichodesmium* proteins of interest (Chapter 3 and 4). In addition, green fluorescent protein (GFP) fusions to protein signal sequences were constructed and their cellular localisation was examined through expression in *Synechocystis* (Chapter 4).

**Deletion mutants:** For the generation of *slr1295* (*futA1*) and *slr0513* (*futA2*) *Synechocystis* deletion mutants (Chapter 4), the antibiotic resistance (Ab\(^R\)) cassette for zeocin (Zeo\(^R\)) from pZEO (Invitrogen, UK) or the chloramphenicol acetyl transferase (*cat*) gene from pACYC184 (NEB, UK), amplified using the primers: *zeo_F* and *zeo_R* or *cat_F* and *cat_R* respectively (Fig. 2.1), were used. PCR products consisting of the sequence upstream of the *futA1/futA2* genes and the beginning of the open reading frame (ORF), or the end of the ORF and downstream flanking DNA, were amplified from *Synechocystis* gDNA using primers *futA1/futA2_ P1* and *P2* or *P3* and *P4* respectively.

![Figure 2.1: Generation of deletion mutants. Using overlap extension (OLE)-PCR, mutagenesis constructs with antibiotic resistance markers (Ab\(^R\)) were designed to](image)

**Figure 2.1:** Generation of deletion mutants. Using overlap extension (OLE)-PCR, mutagenesis constructs with antibiotic resistance markers (Ab\(^R\)) were designed to
homologously recombine to the futA1 or futA2 loci of *Synechocystis* replacing the native genes in mutants ΔfutA1 and ΔfutA2.

Overlap extension (OLE)-PCR (Se. 2.4.2) with primer pair *futA1/futA2* _P1 and P4 was used to generate a mutagenesis construct in which the Ab₈ marker (*ZeoR* and *catR* for ΔfutA1 and ΔfutA2 respectively) was inserted between the upstream and downstream *futA1/futA2* PCR products (Fig. 2.1). This mutagenesis constructs were subsequently used to transform *Synechocystis* (Sec. 2.4.2). Primer *futA1/futA2* _P5, binding internally to *futA1/futA2* was used to test segregation of each mutant.

The ΔfutA1/A2 double mutant was generated by transforming ΔfutA1 with the futA2 gene replacement construct.

**Expression of Trichodesmium genes in Synechocystis:** The function of *Trichodesmium* proteins encoded by genes *Tery_0365-0368* (Chapter 3) and *Tery_3377* (Chapter 4) was examined through expression in *Synechocystis* under the effect of the *psbA2* (*slr1311*) promoter. Gene *psbA2* encodes the D1 protein of photosystem II (PSII) and three isogenes are present in *Synechocystis*, allowing replacement of *psbA2* without any important phenotypic consequences (Mohamed & Jansson 1989).

In the case of *Tery_3377*, the gene (1050 bp) was codon optimised for *Synechocystis* as initial attempts to express the native gene were unsuccessful. The synthetic gene (Integrated DNA Technologies Inc., IA, USA), supplied in plasmid pIDTSMART-AMP, was designed with *NdeI* and *BglII* restriction sites at the 5’ and 3′ ends respectively. The lyophilised plasmid DNA was resuspended in IDTE buffer (10 mM Tris and 0.1 mM EDTA, pH 7.5-8.0) to a concentration of 50 ng/μl. Subsequently, 400 ng plasmid was digested with *NdeI* and *BglII* and the gene sub-cloned into the same sites of the pFLAG vector (Hollingshead et al. 2012; Canniffe et al. 2013). The pFLAG vector recombines into the *Synechocystis* genome such the gene of interest replaces the *psbA2* gene and is therefore expressed under the control of the *psbA2* promoter. The plasmid also confers kanamycin resistance (*kanR*) allowing selection of transformed cells (Fig. 2.2).

For Chapter 4 PCR products comprising the *Trichodesmium* genes *ptxABCD* (*Tery_0365-0368*), *ptxAB* (*Tery_0365-0367*) or *ptxD* (*Tery_0368*) were amplified from *Trichodesmium* gDNA with Q5 polymerase using primers *ptxABCD* _F and *ptxABCD* _R; *ptxABCD* _F and *ptxABCD* _R; or *ptxD* _F and *ptxABCD* _R respectively. All fragments were digested with restriction enzymes *BamH1* and *NdeI* and were ligated into the *BglII* and *NdeI* sites of
pFLAG. All plasmids (Table 2I) were sequence verified (Eurofins MWG Operon, Ebersberg, Germany) and used to transform *Synechocystis*.

**Figure 2.2: pFLAG vector map.** Inserts (light pink) were incorporated between *NdeI* and *BglII* restriction sites adjacent, but transcribed in the opposite direction, to the kan^R^ resistance marker (orange). The *Synechocystis psbA2* us and ds DNA sequence (fuchsia) allows homologous recombination of the insert and kan^R^ at the *psbA2* locus, and sets the gene of interest under the control of the *psbA2* promoter. Plasmid selection in *E. coli* is conferred through ampicillin resistance (amp^R^, orange) Plasmid annotation was performed in PlasmMapper Version 2.0 (Dong et al. 2004).

**Table 2I: Plasmid generated in this study**

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Gene (origin)</th>
<th>Antibiotic Marker</th>
<th>Chapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>pFLAG::ptxABCD</td>
<td>Tery_0365-0368</td>
<td>kan^R^/amp^R^</td>
<td>3</td>
</tr>
<tr>
<td>pFLAG::ptxABC</td>
<td>Tery_0365-0367</td>
<td>kan^R^/amp^R^</td>
<td>3</td>
</tr>
<tr>
<td>pFLAG::ptxD</td>
<td>Tery_0368</td>
<td>kan^R^/amp^R^</td>
<td>3</td>
</tr>
<tr>
<td>pFLAG::3377</td>
<td>Tery_3377</td>
<td>kan^R^/amp^R^</td>
<td>4</td>
</tr>
<tr>
<td>pET21a_3377</td>
<td>Tery_3377</td>
<td>amp^R^</td>
<td>4</td>
</tr>
</tbody>
</table>
**GFP fusions:** The signal sequences of *Synechocystis* FutA1 and FutA2, *Trichodesmium* Tery_3377, and *E. coli* TorA (Chapter 4) were fused in-frame to super folded green fluorescent protein (sfGFP) by OLE-PCR while the cat cassette (cat^R^) was included downstream of GFP to allow selection. The constructs were designed to homologously recombine into the *Synechocystis* genome exactly in place of the *psbA2* ORF (Fig. 2.3).

The DNA sequence corresponding to the predicted (in the case of futA1, futA2, Tery_3377) (Chapter 4, Table 4B) or known (torA; Thomas et al. 2001) signal sequence and first 4 amino acid residues of the mature protein (as identified by Signal P v4.1/v3.0 sensitivity and TatP v1.0 servers, Bendtsen et al. 2005; Petersen et al. 2011) was amplified from the respective organisms gDNA using the corresponding primer pair: futA1_ss_F and futA1_ss_R; futA2_ss-F and futA2_ss_R; 3377_ss_F and 3377_ss_R; or torA_ss_F and torA_ss_R. Approximately 400 bp immediately preceding the *psbA2* gene (amplified with primer pair psbA2_us_F and psbA2_us_R) was joined to each signal sequence in an OLE-PCR (Fig. 2.3) (with primers psbA2_us_F and the respective ss_R primer) such that the ATG start codon of the signal sequence replaced the ATG start codon of *psbA2*. The GFP coding sequence was isolated from plasmid pED151 (provided by A.Hitchcock) with primers GFPss_F and GFPss_R and ~400 bp of sequence immediately downstream of *psbA2* was amplified using the primer pair psbA2_ds_F/psbA2_ds_R. The cat^R^ marker was inserted between the GFP and downstream *psbA2* flank by OLE-PCR (Fig. 2.3) with the extreme primers GFP_F and psbA2_ds_R. Finally, each *psbA2* upstream-signal sequence fragment was joined to the GFP-cat-psbA2 downstream fragment using OLE-PCR (Fig. 2.3) with primers psbA2_us_F and psbA2_ds_R.

To create GFP fusions to the twin arginine mutated signal sequence of Tery_3377 (Tery_3377ss-KK-GFP) primer 3377_ss_F was switched for 3377-R5/6K-GFP and psbA2_us_R for psbA2-us-R2. These alternative primers were designed to align during OLE-PCR and both contained the first 18 bp of the Tery_3377 signal sequence with changed nucleotide sequence between positions 13-18 (AAA AAA in the place of AGA CGA) such that the translated sequence codes for two lysine residues in the place of the two arginine residues of the twin arginine transport (TAT) system. Following the construction of the GFP tagged signal sequence fragments and after complete segregation of each strain modified DNA segments were sent for sequencing (Eurofins MWG Operon, Ebersberg, Germany) to assure the correct sequence identity of each.
Figure 2.3: Generation of the GFP tagged signal sequence strains. (A) DNA fragments were amplified and sequentially joined through OLE-PCR to assemble the GFP tagged signal sequence (ss) constructs. (B) As these were bordered by the psbA2 us and ds regions, transformations and selection on chloramphenicol (resistance conferred with cat<sup>R</sup>) resulted in psbA2 replacement placing the construct under the control of the psbA2 promoter.

**Protein over-expression in E. coli:** To express C-terminally hexa histidine-tagged Tery_3377 in *E. coli*, the ORF (omitting the predicted N-terminal signal peptide and stop codon) was amplified from the *Trichodesmium* genome using primers 3377_OE_F and 3377_OE_R. The resulting PCR product was cloned into the Ndel/Xhol sites of pET21a(+) (Novagen, Madison, WI) generating the plasmid pET21a_3377.
2.4.2 Techniques

Overlap extension (OLE)-PCR was used to assemble fragments of amplified DNA into larger linear constructs which were used to transform Synechocystis in Chapter 4.

For OLE-PCRs Accuzyme Polymerase Mix (Bioline, London, UK) or Q5 High-Fidelity Polymerase Mix (New England Biolabs Ltd, Hitchin, UK) were used. Oligonucleotide primer pairs (Table A1) constructed as detailed earlier (Sec. 2.3.3) were diluted to 100 nM final, reaction concentration. They were used to amplify fragments from gDNA or plasmid DNA templates. Primers were designed to attach to ~15-25 bp of the 3’ or 5’ end of the fragment being amplified and generate ~ 25bp overhangs with complementarity to the neighbouring fragments (overlapping regions).

Constructs were then assembled by connecting the fragments using flanking primers at: 98 °C for 30 secs (initial denaturation), 30 cycles of 98 °C for 10 secs (denaturation), 30 secs of annealing at 55- 60 °C, and 72 °C for the required extension time (30 secs for each 500 bp or 1000 bp for Accuzyme and Q5 polymerase respectively) and a final extension at 72 °C for 10 mins. The primers were added to the reaction after the first 5 of 30 PCR cycles.

All PCR generated fragments and constructs were visualised by gel electrophoresis (Sec. 2.3.4) to verify their correct size. Amplified fragments and constructs were purified using the QIAqucik PCR Purification Kit if a single product was identifiable or the QIAqucik Gel Extraction Kit (Qiagen, Manchester, UK) to separate the desired product from non-specific bands/plasmid template DNA.

Restriction enzyme digestions and dephosphorylation of 5’-ends: Digestions were performed with 1 μl NdeI, BglII, AfeI and MfeI restriction enzymes (Thermo Scientific, DE, USA), 0.6-1 μg vector or 0.1-0.15 μg insert according to the manufacturer’s instructions. To remove the 5´ phosphoryl termini and prevent recircularization of cloning vectors they were treated with Antarctic Phosphatase (New England Biolabs Ltd, Hitchin, UK) for 30 mins at 37 °C. Restriction enzyme digested DNA and Antarctic Phosphatase reactions were QIAqucik purified (PCR Purification Kit, QIAGEN, Manchester, UK) prior to ligations.

Ligations and purifications: Digested vectors and fragments (1:3 -1:5 vector to insert ratio) were ligated by incubating with 1 μl T4 DNA Ligase (New England Biolabs Ltd, Hitchin, UK) for 2 hrs or overnight at room temperature according to the manufacture’s protocol. Reactions were purified by precipitating the ligated DNA with n-butanol, removing the
supernatant (centrifugation $12 \times 10^3$ rpm for 10 mins) and drying the pellet (vacuum centrifuge, 2 mins at 45 °C).

**Transformations:**

**E. coli:** Plasmids were transformed into *E. coli* cloning strains through heat shock (DH5-α, BL21) or electroporation (XL1-blue).

Chemically competent DH5-α and BL21 cells (New England Biolabs Ltd, Hitchin, UK) were mixed with the ligated plasmids. The cell mixture was left on ice for 30 mins and then transferred to a 42 °C water bath for 30 secs after which it was cooled again on ice for 2 mins. Super Optimal Broth (SOC) (1 ml) was added to the cells which were incubated at 37 °C shaking at 250 rpm for 60 mins before plating 200 μl on LB-ampicillin (100 μg/ml) plates which were incubated overnight at the same temperature.

For electroporation XL1-blue competent cells were defrosted on ice, mixed with ligated plasmids and incubated on ice for 60 secs. After that, the cell/plasmid mix was transferred to a cold electroporation cuvette for electroporation. They were then immediately supplemented with 1 ml SOC and incubated at 37 °C shaking at 250 rpm for 60 mins. Cells were then centrifuged ($8 \times 10^3$ rpm for 2 mins), resuspended in 200 μl and plated on LB-ampicillin plates to be incubated overnight at 37 °C.

Colonies were screened by colony PCR with appropriate primers and plasmids were prepped from successful clones using the QIAprep Spin Miniprep Kit (Qiagen, Manchester, UK) according to the manufacturer’s instructions.

**Synechocystis:** Wild-type (WT) *Synechocystis* was transformed as described by Williams, (1988). Plasmid pFLAG or linear mutagenesis fragments (as described above) were incubated with 200 μl of cell culture at 30 °C for 1 hr and recovery was allowed for 24 hours on a BG-11 plate with no antibiotic. Genome copies were segregated by repeated streaking on BG-11 plates with increasing antibiotic concentration and transformants homozygous for the deletions were verified by PCR using template DNA derived from WT or transformed cells.

The correct DNA sequence in the resulting segregated strains was confirmed through automated DNA sequencing (Eurofins MWG Operon, Ebersberg, Germany).
2.5 Protein Immunoblotting and Crystallisation

2.5.1 Immunoblotting

Proteins were extracted from *Synechocystis* and *E. coli* through suspension of pelleted cells in 100 μl of 10 mM Tris-HCl pH 7.3 and heating for 5 mins at 100 °C followed by centrifugation at 13 x10^3 rpm for 5 mins at 4 °C. *Trichodesmium* protein extraction were achieved through 3, 1 min sonication cycles, the first two followed by freezing in liquid nitrogen and the third by centrifugation for 5 mins at 13 x10^3 rpm at 4 °C (Brown et al. 2008).

Protein concentrations were assessed using the Pierce™ BCA Protein Assay Kit (Thermo Scientific, DE, USA) with bovine serum albumin (BSA) standards according to the manufacturer’s instructions.

Immunoblotting was performed as described by Brown et al. (2008). Samples were incubated at 80 °C for 5mins with 50 mM dithiothreitol (DTT) and run on 4–12% NuPAGE acrylamide gradient gels (Invitrogen, Paisley, UK) in MES running buffer (Invitrogen, Paisley, UK).

Primary antibodies anti-IdiA and anti-Slr0513 (generously provided by E. Pistorius) and anti-GFP (Sigma-Aldrich, St. Louis, USA) were diluted to 1:500, 1:1000 and 1:5000 respectively in 2% ECL Advance blocking reagent (GE Healthcare Little Chalfont, UK) Tris-buffered saline plus Tween (TBS-T). Dilutions of 1:10000 of horseradish peroxidase-conjugated anti-rabbit secondary antibody in 2% blocking agent TBS-T (Abcam, Cambridge, United Kingdom) were used with all primary antibodies. Primary and secondary antibody incubations were performed at room temperature for 1 hr.

For imaging a VersaDocTM CCD imager (Bio-Rad Laboratories Ltd, Hemel Hempstead, UK) was used after application of ECL Advance detection reagent (GE Healthcare, Little Chalfont, UK) as in Richier et al. (2012).

2.5.2 Purification and crystallisation

The *Trichodesmium* protein Tery_3377 (Chapter 4) was purified following over production in *E. coli* BL21(DE3) and used to set up crystal trials as a first step of resolving its three dimensional structure by X-ray crystallography.
Cells grown from a single colony on an agar plate were used to inoculate 100 ml LB-ampicillin (100 μg/ml) and were incubated at 37 °C shaking at 200 rpm overnight. The following day 10 ml was used to inoculate four 1 L batches of LB-ampicillin which were maintained at the same conditions until OD<sub>600</sub> was equal to 0.6. All four flasks were then induced with Isopropyl β-D-1-thiogalactopyranoside (IPTG) at 1 mM final concentration and two were placed back at 37 °C for 3 hrs. The remaining two were left shaking at 18 °C overnight.

Cells were harvested by centrifugation at 5 x 10<sup>3</sup> rpm for 20 mins and the pellet was transferred to 50 ml PBS to be centrifuged again for 20 mins at 4 x 10<sup>3</sup> rpm. The supernatant was removed and cell pellets were either stored at -20 °C or processed directly for protein extraction. Cells were resuspended in Lysis buffer (Table 2J) and protein was extracted in an ice-water bath by sonication for a total duration of 5 mins (10/30 secs on/off cycles). The lysate was subsequently centrifuged at 45 x 10<sup>3</sup> rpm for 30 mins at 4 °C.

Purification of His-Tagged Tery_3377 was achieved using a Ni-column for immobilised metal affinity chromatography (IMAC) and a subsequent size exclusion chromatography (SEC) step. The Ni-column (4 ml Ni-NTA Superflow resin, Qiagen, Manchester, UK) was equilibrated with Binding Buffer (Table 2J) before applying the lysate which was previously filtered through a 2 μm filter. The lysate was run through the column 5-6 times after which the Wash buffer (Table 2J) was applied followed by the Elution buffer (Table 2J). Subsequently, the protein was concentrated to less than 1 ml with spin concentrators (Generon Ltd., Berkshire, UK) and applied to the HiLoad 16/60 Superdex 200 SEC column (GE Healthcare Life Sciences, Little Chalfont, UK) which was equilibrated with SEC Buffer (Table 2J). The fractions that contained Tery_3377 (as determined by SDS-PAGE and Coomassie-staining) were pooled together and the protein was spin concentrated to a volume of less than 100 μl and a concentration of 41.620 mg/ml.

For crystallisation trials the two 96 condition, 3D protein crystallization screens Morpheus and JBSC were set in sitting drop 96-well plates (INTELLI-PLATE) using a Gryphon liquid handling robot (ARI, CA, USA).
Table 2J: Buffers used for Tery_3377 protein extraction

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Recipe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lysis</td>
<td>50 mM Tris, pH 8.5 (1 pH below pI of Protein) 300 mM NaCl 5% Glycerol 2 mM β-Mercaptoethanol (fresh on day of use)</td>
</tr>
<tr>
<td>Binding</td>
<td>Same as Lysis 20 mM Imidazole</td>
</tr>
<tr>
<td>Wash</td>
<td>Same as Binding 50 mM Imidazole</td>
</tr>
<tr>
<td>Elution</td>
<td>Same as Binding 150-300 mM Imidazole</td>
</tr>
<tr>
<td>SEC</td>
<td>50 mM Tris, pH 8.5 150 mM NaCl</td>
</tr>
</tbody>
</table>

2.6 RNA sequencing

The transcriptomic response of *Trichodesmium* grown under different Fe regimes (Fe+, Fe-, [Dust] and Dust+) (as described in Chapter 5) was analysed by next generation sequencing at the National Oceanography Centre Sequencing Facility.

2.6.1 Library preparation and sequencing

Complementary (c)DNA libraries were prepared using the TruSeq Stranded mRNA Library Prep Kit (Illumina), which was modified for use with bacterial RNA samples by replacing the initial poly-A-based mRNA isolation step with probe-based depletion of ribosomal (rRNA). Subsequently, rRNA transcripts were selectively depleted from the RNA samples using a bacterial Ribo-Zero rRNA Removal Kit (Illumina INC., Ca, USA), according to the manufacturers’ instructions. The precipitated RNA pellets were resuspended directly in 18 μl of Fragment-Prime-Finish mix from the TruSeq Stranded mRNA kit, and RNA fragmentation, cDNA synthesis and library construction were carried out as described in the kit protocol.

For sequencing the molarity of each library was determined from the measured concentration and average fragment size using equation 3.

Equation 3:

\[
\text{conc (nM)} = \frac{\text{conc (ng}\mu\text{l}^{-1})}{\text{average fragment size} \times 660 \text{ g mol}^{-1} \times 10^6}
\]
Libraries were then diluted to a concentration of 10 nM and pooled before denaturation to single strands. The combined sequencing library was sequenced on an Illumina MiSeq instrument using paired-end sequencing, a read length of 151 bp, and a 6 bp index read.

Image analysis, base calling and de-multiplexing of barcoded samples was carried out on-instrument (MiSeq Control Software). Raw sequencing reads were obtained in fastq format. Overall sequencing quality was assessed (FastQC version 0.11.3, Babraham Institute, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) before fastq files were processed (CutAdapt v1.8.1, http://code.google.com/p/cutadapt/) to remove TruSeq adapter sequences and low quality bases from the 3’ end of reads.

The trimmed reads were mapped against the *Trichodesmium erythraeum* IMS101 reference genome assembly (Ensembl Bacteria database, accession GCA_000014265.1) using the TopHat spliced read aligner v2.0.14 (Kim et al. 2013). For each RNA sample, the number of reads mapping to each annotated gene (Ensembl Bacteria database, accession GCA_000014265.1) in IMS101 were counted (HTSeq-count script, HTSeq framework, version 0.6.0, (Anders et al. 2015)). Mapped reads were sorted by read name (Samtools version 1.2 (Li et al. 2009)), and reads mapping to each feature annotated as ‘gene’ were counted. Reads were processed using the -stranded=reverse option, to correctly account for strand-specific read mapping.

### 2.6.2 Statistical analysis

Differential gene expression analysis was carried out using the Bioconductor package DESeq2 version 1.8.2 (Love et al. 2014), running on R version 3.2.0. using an ANODEV approach implemented in the DESeq package (Anders & Huber 2010). Likelihood ratio testing was used to compare the fit of read count data against a negative binomial model in a one-factor experimental design where replicates were grouped by growth condition, against a reduced model where condition information was removed. Pairwise contrasts between conditions were calculated using a Wald test. Benjamini-Hochberg correction for multiple testing was used and genes with a p value of < 0.05 were classed as significantly differentially expressed either for differentially expressed genes across the 4 treatments or between specific pairs of treatments (pairwise comparisons).
2.7 Bioinformatic analysis

2.7.1 Analysis of the Dust-FeCl$_3$ *Trichodesmium* transcriptome

In Chapter 5 expression of the identified differentially expressed genes from the Fe+, Fe-, [Dust] and Dust+ cultures (3 biological replicates for each) was clustered and a dendrogram was constructed (Minitab 17.3.1, Minitab Inc., Coventry, UK). Distances were calculated using correlation coefficients and amalgamated using average linkage.

Grouping and categorizing of the gene ontology (GO) classes (downloaded from CyanoBase http://genome.microbedb.jp/cyanobase/TERY) of the differentially expressed genes was performed using a generic GO Slim (GO_Slim) for which counts were accumulated (CateGOrizer, Hu et al. 2008). Enrichment of GO classes in groups of differentially expressed genes against all expressed genes identified in the transcriptome was completed using Singular Enrichment Analysis (AGRIGO, Du et al. 2010).

Further analysis of uncharacterised proteins utilised information on conserved protein domains (NCBI Conserved Domain Database, Marchler-Bauer et al. 2015), subcellular localisation or metal binding sites (UniProtKB Automatic Annotation pipeline Consortium 2015). Signal sequence prediction software Signal P v4.1 (v3.0 sensitivity) and TatP v1.0 servers (Bendtsen et al. 2005; Petersen et al. 2011) were also used.

To visualise the pattern of expression along gene clusters and unannotated or suspected miss-annotated regions, sample reads aligned against the *Trichodesmium* genome were inspected manually through visualisation in the Artemis genome browser, (Rutherford et al. 2000).

2.7.2 Investigation of omic datasets from other studies

Conservation of *Trichodesmium* phosphite utilisation genes (Chapter 4) *in situ* was examined by searching for homologues of *ptxABCD* (*Tery_0365-0368*) in three published genomic datasets, consisting of two draft genomes (*T. erythraeum* 21-75 and *T. theibautii* H9-4) and an environmental metagenome from the Atlantic Ocean (Walworth et al. 2015). Searches were performed using tBlastn (Altschul et al. 1990) with an e-value cut-off of 0.000001.

To investigate expression of genes *ptxABCD* (*Tery_0365-0368*) in transcriptomic datasets, four publically available transcriptomic datasets were searched. The datasets consisted of...
two *in situ* meta-transcriptomes, one from the Amazon River Plume (Hilton et al., 2014) and one from the South Pacific Subtropical Gyre (Hewson et al. 2009), and two transcriptomes from cultured samples (Pfreundt et al., 2014; unpublished dataset: NCBI accession PRJNA237745). Illumina datasets were searched for reads aligning to *Tery_0365-0368* with > 90% sequence identity over 100% of the read length using MegaBlast (Morgulis et al. 2008), via the SRA Blast service from NCBI. 454 datasets were searched for reads encoding homologues of *Tery_0365-0368* using tBlastn with an e-value cut off of 0.000001 and a minimum percentage identity of 90%. Expression of genes *phnDCEEGHIJKLM* (*Tery_4993-5003*) in the cultured transcriptomes (Pfreundt et al., 2014; unpublished dataset, NCBI accession PRJNA237745) was assessed in a similar fashion.

### 2.8 Epifluorescent and confocal microscopy

GFP expressing *Synechocystis* strains were viewed using an axioscope Z plus epifluorescence microscope (Zeiss, Oberkochen, Germany) and photographed with an HRm axioCam (Zeiss, Oberkochen, Germany). Samples where excited with 400 – 490 nm light filtered with a FITC filter set.

Strains were also visualised under the Leica SP5 LSCM (Leica Microsystems, Wetzlar, Germany) confocal microscope. Cells were exposed to high light for 15 mins prior to visualisation to increase the expression of the GFP fusions from the *psbA2* promoter. As described in Aldridge et al. (2008) excitation was performed through a 488 nm laser line at which GFP emits between 510 to 530 nm while phycocyanin, used to represent pigment autofluorescence, emits between 600 and 700 nm.

### 2.8 Phylogenetic analysis

Maximum likelihood phylogenetic analysis of *Tery_3377* homologues from *Synechocystis* and selected bacteria was performed to identify possible functional clusters (Chapter 4).

The *Trichodesmium* *Tery_3377* and two *Synechocystis* proteins FutA1 and FutA2 with their closest homologues from *Lyngyba sp. PCC 8106*, *Cyanothece sp. CCY0110* and *Phormidium willei* respectively (identified using BlastP, Altschul et al. 1990), as well as homologues from *Prochlorococcus marinus* MIT 9301, *Synechococcus elongatus* PCC 6301 and PCC 7942,
Anabaena sp. PCC 7120, *Haemophilus influenza* and *Neisseria meningitides* were included in the analysis (Table 2K).

**Table 2K: Tery_3377 homologues included in the phylogenetic analysis**

<table>
<thead>
<tr>
<th>Organisms</th>
<th>Protein identifier</th>
<th>Size (AA)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Trichodesmium erythraeum</em> IMS101</td>
<td>Tery_3377</td>
<td>349</td>
</tr>
<tr>
<td><em>Synechocystis</em> PCC 6803</td>
<td>Slr1295 (FutA1)</td>
<td>360</td>
</tr>
<tr>
<td></td>
<td>Slr0513 (FutA2)</td>
<td>346</td>
</tr>
<tr>
<td><em>Lyngbya</em> sp. PCC 8106</td>
<td>WP_009786151.1</td>
<td>350</td>
</tr>
<tr>
<td><em>Cyanothece</em> sp. CCY0110</td>
<td>WP_035799093.1</td>
<td>365</td>
</tr>
<tr>
<td><em>Phormidium</em> <em>willei</em> BDU 130791</td>
<td>OAB59261</td>
<td>346</td>
</tr>
<tr>
<td><em>Prochlorococcus</em> <em>marinus</em></td>
<td>P9301_13611 (AfuA)</td>
<td>340</td>
</tr>
<tr>
<td><em>Synechococcus</em> <em>elongatus</em> PCC 6301</td>
<td>Syc0146_c</td>
<td>340</td>
</tr>
<tr>
<td></td>
<td>Syc1920_d</td>
<td>366</td>
</tr>
<tr>
<td><em>Synechococcus</em> <em>elongatus</em> PCC 7942</td>
<td>Synpcc7942_1409</td>
<td>340</td>
</tr>
<tr>
<td></td>
<td>Synpcc7942_2175</td>
<td>366</td>
</tr>
<tr>
<td><em>Anabaena</em> sp. PCC 7120</td>
<td>Alr1382</td>
<td>336</td>
</tr>
<tr>
<td><em>Haemophilus</em> <em>influenza</em></td>
<td>FbpA</td>
<td>332</td>
</tr>
<tr>
<td><em>Neisseria</em> <em>meningitides</em></td>
<td>FbpA</td>
<td>331</td>
</tr>
</tbody>
</table>

The online platform Phylogeny.fr (Dereeper et al. 2008; Dereeper et al. 2010) was utilised to create an analysis pipeline whereby sequences were firstly aligned (T-coffee, v11.00.8cbe486, Notredame et al. 2000) and curated (G-blocks v0.91b, Castresana 2000). Subsequently, a maximum likelihood phylogenetic tree was reconstructed (PhyML v3.1/3.0, Guindon & Gascuel 2003) and approximate likelihood ratio test was carried out to compute branch supports (aLRT, Anisimova & Gascuel 2006). Finally, a graphical representation was produced (TreeDyn v198.3, Chevenet et al. 2006).
Chapter 3
Phosphite utilization by the globally important marine diazotroph *Trichodesmium*

Parts of this Chapter have been published in:

3.1 Introduction

*Trichodesmium*’s growth and nitrogen (N$_2$) fixing activity appear to be limited by the availability of phosphorus (P) in various oceanic locations (Sañudo-Wilhelmy et al. 2001; Mills et al. 2004; Moutin et al. 2005). A better understanding of the available P sources is necessary to predict restrictions on the organism’s biogeography, both presently and in the future oceans. At the same time, it is crucial to recognise the interdependence between life and the environment that surrounds it as the major marine primary producers themselves (including *Trichodesmium*) considerably affect the oceanic P cycle.
There is emerging evidence that in addition to phosphate (PO$_4^{3-}$), both inorganic and organically complexed, P is present in the marine environment in reduced forms including organophosphonates (Pn) and phosphite (PO$_3^{3-}$) (Table 3A). The concentrations of these reduced forms in the ocean are currently largely unknown (Karl 2014). Studies using $^{31}$P-NMR reported Pn can comprise 25% of the dissolved organic P (DOP) pool (Clark et al. 1998; Kolowith et al. 2001). Further, a recent analysis, using high-performance liquid chromatography coupled to inductively coupled plasma mass spectrometry detection (HPLC-ICP-MS), of freshwater samples has indicated 25% of dissolved P to be in the +3 or +1 oxidation state as PO$_3^{3-}$ and hypophosphite (PO$_2^{3-}$) (Pasek et al. 2014). Also, ion exchange chromatography has been previously used to identify PO$_3^{3-}$ and PO$_2^{3-}$ in geothermal hot springs (Pech et al. 2009; 2011) and lake interstitial water (Han et al. 2013). However, no data is available regarding the concentrations of these sources of P in the ocean.

**Table 3A:** The different forms and oxidation states of phosphorus in the ocean. Inorganic phosphorus can be found as phosphate, phosphite and hypophosphite while organic phosphorus is encountered as organophosph(on)ates.

<table>
<thead>
<tr>
<th>Complexation</th>
<th>Oxidation state</th>
<th>P form</th>
<th>Enzyme for oxidation/hydrolysis</th>
<th>In Trichodesmium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inorganic</td>
<td>+5</td>
<td>Phosphate (PO$_4^{3-}$)</td>
<td>N/A</td>
<td>(Orchard et al. 2003)</td>
</tr>
<tr>
<td></td>
<td>+3</td>
<td>Phosphite (PO$_3^{3-}$)</td>
<td>PtxD: NAD:phosphite oxidoreductase</td>
<td>Not studied</td>
</tr>
<tr>
<td></td>
<td>+1</td>
<td>Hypophosphite (PO$_2^{3-}$)</td>
<td>HtxA: hypophosphite dioxygenase</td>
<td>Unknown</td>
</tr>
<tr>
<td>Organic</td>
<td>+5</td>
<td>Ester-linked organophosphates (C-O-P bond)</td>
<td>APase</td>
<td>PO$_4^{3-}$ regulation (Orchard et al. 2009)</td>
</tr>
<tr>
<td></td>
<td>+3</td>
<td>Organophosphonates (Pn) (C-P bond)</td>
<td>C-P lyase</td>
<td>PO$_4^{3-}$ regulation (Dyhrman et al. 2006)</td>
</tr>
</tbody>
</table>

*Trichodesmium* has the ability to process Pn (Dyhrman et al. 2006), and appears to have a significant role in its production (Dyhrman et al. 2009) as Pn can make up 10% of cellular particulate organic P (POP). In addition, potential uptake of PO$_3^{3-}$ by *Trichodesmium* colonies *in situ* was recently suggested (Van Mooy et al. 2015). If PO$_3^{3-}$ is indeed available to
**Chapter 3: Phosphite utilisation by *Trichodesmium***

*Trichodesmium*, and the observations are not the mere effect of uptake by other members of the *Trichodesmium* colony consortia or cell surface adsorption, a previously unrecognised pool of P could be available to fuel the organism’s growth. The ability to process PO$_3^{3-}$ could also equate to more ‘bridges’ interconnecting P species in a complex intracellular P redox network. Consequently, potential synthesis of PO$_3^{3-}$ as an intermediate or end product in this network should come as no surprise (Van Mooy et al. 2015).

Although both C-P lyase (Metcalf & Wanner 1991), and APase (Yang & Metcalf 2004) (enzymes involved in Pn and organophosphate catabolism respectively) may have a secondary activity in driving the oxidation of PO$_3^{3-}$ to PO$_4^{3-}$ (Metcalf & Wanner 1991; Yang & Metcalf 2004), in 1998, Metcalf & Wolfe found evidence of a specific PO$_3^{3-}$ oxidising enzyme which was later characterized as phosphite dehydrogenase (PtxD) (Costas et al. 2001). This enzyme couples PO$_4^{3-}$ production to NADH formation. The release of NADH potentially makes PO$_3^{3-}$ oxidation more energetically favourable and could determine PtxD as an enzyme of double functionality converting PO$_3^{3-}$ to PO$_4^{3-}$ and producing energy in the process. For example, for the sulphate reducing *Desulfitotignum phosphitoxidans*, PO$_3^{3-}$ oxidation provides electrons to drive the organism’s energy metabolism (Schink & Friedrich 2000; Schink et al. 2002). Due to its ability to reduce NAD, PtxD has been targeted as an enzyme of interest in the regeneration of NADH (reductant) in industrial applications (Hirota et al. 2012).

Phosphite dehydrogenase (*ptxD*) was first identified to be part of a 5 gene PO$_3^{3-}$ utilisation operon (*ptxABCD*) in *Pseudomonas stutzeri* WM88, with *ptxABC* encoding the members of an ATP-binding cassette (ABC) transporter (an ATPase, periplasmic binding protein and permease respectively), followed by *ptxE*, which encodes a putative transcriptional regulator (Metcalf & Wolfe 1998). More recently, homologs of all the genes apart from *ptxE* were identified and characterised in the marine species *Prochlorococcus* MIT9301 (Martínez et al. 2012). The authors also indicated the presence of *ptx* gene homologs in other marine bacteria including diazotrophic *Cyanothece*, *Nodularia* and *Trichodesmium*.

This Chapter attempts to identify whether *Trichodesmium erythraeum* IMS101 in culture can be sustained with PO$_3^{3-}$ as the only source of P. Further, the role of the putative *ptxABCD* genes and the conditions controlling their expression are described. Due to the inability to directly study the function of PtxABCD in *Trichodesmium* through deletion mutants (as molecular genetic manipulation techniques are lacking in this organism) we use heterologous expression in the model freshwater cyanobacterial species *Synechocystis* PCC.
6803 to demonstrate their function in PO$_3^{3-}$ utilisation. The environmental relevance of this work is discussed and the prevalence of the four genes/proteins in published environmental omic studies is reported.

### 3.2 Results and discussion

#### 3.2.1 Phosphite as a source of P for *Trichodesmium*

To determine if PO$_3^{3-}$ can support the growth of *Trichodesmium*, cultures of *T. erythraeum* IMS101 (*Trichodesmium*) in YBC-II media with no P additions were supplemented with 50 μM PO$_3^{3-}$ (Na$_2$HPO$_3$.5H$_2$O) and their growth was compared to those provided with equivalent concentrations of PO$_4^{3-}$ (50 μM NaH$_2$PO$_4$.H$_2$O) and P-deficient cultures (no P added) for 13 days (Fig. 3.1).

Maximum growth rates measured in the presence of PO$_4^{3-}$ under our laboratory conditions (0.2 day$^{-1}$), were 29% higher but not significantly different to the PO$_3^{3-}$ treatment (0.14 day$^{-1}$). P-starved cultures (P-) grew slower (Student’s t-test, P < 0.05) than both of the other treatments (Fig. 3.1B) with biomass declining after day 7 of the experiment and establishing a significantly lower number of cells (Student’s t-test, P < 0.01) from day 10 onwards (Fig. 3.1A). This indicates that the P-omitting control-treatment did reach P starvation after a few days, with survival and minimal growth during the first week possibly maintained through P carryover during inoculation of the experiment. Although the final cell concentration is lower in the PO$_3^{3-}$ compared to the PO$_4^{3-}$ treatment (Student’s t-test, P < 0.05) (Fig.3.1A), a result comparable to that reported for *Prochlorococcus* MIT9301 (Martínez et al. 2012), these data clearly demonstrate that *Trichodesmium* has the capability to utilize PO$_3^{3-}$ as its sole source of P.

The reduced growth on PO$_3^{3-}$ compared to PO$_4^{3-}$ possibly indicates an insufficient P supply from the former due to a lower efficiency of its uptake pathway. Another hypothesis could be that a higher PO$_3^{3-}$ to PO$_4^{3-}$ ratio, which can be caused by preferential PO$_4^{3-}$ consumption/depletion, is a stimulus that downregulates cellular growth rates as part of the cells’ response strategies to low P. In situ, maintaining a flexible growth rate is part of a trophic strategy that can confer longer lasting growth on the limited nutrient reserves of starved oceans maintaining survival until encounter with new nutrient inputs (Lauro et al. 2009; Yooseph et al. 2010; Kirchman 2015). Therefore, this strategy is potentially advantageous in fluctuating conditions such as those encountered during atmospheric
Chapter 3: Phosphite utilisation by Trichodesmium

Figure 3.1: Growth and physiology of Trichodesmium in different P regimes. Cultures were added phosphate (PO₄³⁻, dark blue triangles), phosphite (PO₃³⁻, cyan squares) or no P (P⁻, orange circles) and cell concentrations were measured over 13 days. Asterisks (**) indicate significant difference in cell numbers (Student’s t-test, P < 0.01) to other treatments on the same day (A). Growth rates were calculated for the three treatments during exponential growth (B) and filament length (C) along with the number of cells per filament (D) were computed for the last day of the experiment. Letter groups (x, y and z) are significantly different to one another (Student’s t-test, P < 0.05). Error bars represent the standard deviation from the mean of three biological replicates.
deposition or upwelling events occurring in the oligotrophic oceans where *Trichodesmium* is abundant. This strategy, referred to as a ‘feast or famine’ lifestyle, was previously linked to marine microorganisms with larger genome sizes (Lauro et al. 2009; Yooseph et al. 2010; Kirchman 2015) as seen in *Trichodesmium*.

In the absence of PO$_4^{3-}$ a difference in the morphology of the filaments is also observed (Fig. 3.1C/D). On the last day of the experiment, filaments growing with PO$_4^{3-}$ were approximately half the length and comprised two times less cells (Student’s t-test, P < 0.01) compared to those with PO$_4^{3-}$. Although filaments seemed to also be shorter in PO$_3^{3-}$ compared to without P (Student’s t-test, P < 0.01) the number of cells per filament in the two treatments is comparable due to a small difference in average cell length (not statistically significant-data not shown). The difference in filament morphology across the experimental treatments could therefore be explained by two independent factors acting in the opposite direction: (a) sensing of decreased PO$_4^{3-}$ in the immediate environment of the cells (in PO$_3^{3-}$ and P-) potentially stimulates a response to decrease the number of cells per filament and therefore increase the surface area to volume ratio and the uptake of P, while (b) P-stress (in P-) reduces cell division rate resulting in longer cells and longer overall filaments.

### 3.2.2 Description of the *Trichodesmium* putative *ptx* genes

Homologues to the *Prochlorococcus* MIT9301 PO$_4^{3-}$ utilisation *ptxABCD* genes with similarities (64-70%) are present in the *Trichodesmium* genome (*Tery_0365-0368*). These are annotated as a Pn transporter (*Tery_0365-0367*) and an NAD binding dehydrogenase (*Tery_0368*) (Table 3B). An intragenic space of 449 bp separates *Tery_0365* from the remaining three genes (Fig. 3.2). No homologue to the transcriptional regulator *ptxE* found in *Pseudomonas stutzeri* WM88 is identified. The *Tery_0365-0368* genes will be referred to as *ptxABCD* hereafter and reasons for the reannotation are provided through this Chapter.

![Figure 3.2: The arrangement of the *ptxABCD* genes in *Trichodesmium*.](image.png)

The *ptxA* (759 bp) *ptxB* (897 bp) and *ptxC* (819 bp) genes encode members of an ATP binding cassette (ABC) transporter and are located directly upstream of the NAD-dependant dehydrogenase (*ptxD*; 996 bp).
Table 3B: Details of the ptx homologous genes in *Trichodesmium*

<table>
<thead>
<tr>
<th>Putative gene</th>
<th>Tery number</th>
<th>Annotation</th>
<th>Putative function</th>
<th>Nucleotide identity to Prochlorococcus str. MIT 9301</th>
</tr>
</thead>
<tbody>
<tr>
<td>ptxA</td>
<td>0365</td>
<td>phosphonate ABC transporter, ATPase subunit</td>
<td>ATPase of ABC transporter</td>
<td>70%</td>
</tr>
<tr>
<td>ptxB</td>
<td>0366</td>
<td>phosphonate ABC transporter, periplasmic phosphonate-binding protein</td>
<td>Periplasmic binding protein of ABC transporter</td>
<td>67%</td>
</tr>
<tr>
<td>ptxC</td>
<td>0367</td>
<td>phosphonate ABC transporter, inner membrane subunit</td>
<td>Permease of ABC transporter</td>
<td>69%</td>
</tr>
<tr>
<td>ptxD</td>
<td>0368</td>
<td>D-isomer specific 2-hydroxyacid dehydrogenase, NAD-binding</td>
<td>Phosphite dehydrogenase</td>
<td>64%</td>
</tr>
</tbody>
</table>

3.2.3 Regulation of ptx gene expression in *Trichodesmium*

RNA extracted on the final day of the growth experiment (Fig. 3.1) was used for gene expression analysis by quantitative polymerase chain reaction (qPCR) to investigate whether *ptxB* (representative of the ABC-transporter) and *ptxD* were regulated in response to the nature of P source. Transcript levels were corrected using reference genes *rnpB* (Chappell & Webb 2010) and *rotA* (Orchard et al. 2009). A significant increase (one-way ANOVA, *P* < 0.05) in expression of both genes (1.5–2.0 relative fold change) was identified when growing with PO\(_3^-\) or no P compared to PO\(_4^{3-}\) (Fig. 3.3). Although relatively small, the observed upregulation of genes possibly involved in PO\(_3^-\) uptake and oxidation suggests PO\(_3^-\) utilization to be a dynamic process responding to environmental PO\(_4^{3-}\)-limitation and/or the presence of PO\(_3^-\).

The fact that the genes are expressed under P-replete conditions at a basal level is also confirmed through identification in published (Pfreundt et al. 2014) and unpublished (NCBI accession: PRJNA237745, dataset described in Chapter 5) transcriptomes of *Trichodesmium* in culture (Table 3C). To compare, the same transcriptomic studies were also scanned for genes *phnDCEEGHJKLM* (*Tery_4993-5003*) involved in Pn metabolism. These were previously reported to only be expressed under P-limitation in the environment (Dyhrman et al. 2006) but their expression is identified in transcriptomic studies of P-replete...
Trichodesmium in culture. Therefore, although upregulation of both $\text{PO}_3^{3-}$ and Pn metabolism genes occurs under $\text{PO}_4^{3-}$-limitation, a basal level of expression seems to be constantly maintained. This is consistent with the hypothesis that dynamic cycling of P between valance states occurs inside Trichodesmium itself, and possibly within/between members of the consortia in situ (Van Mooy et al. 2015).

![Figure 3.3: Gene expression analysis in response to P source.](image)

Expression of the periplasmic substrate-binding transporter component ($\text{ptxB}$) and phosphite dehydrogenase ($\text{ptxD}$) was quantified using qPCR and normalized against the reference genes $\text{rnpB}$ and $\text{rotA}$. Expression in cultures with phosphite ($\text{PO}_3^{3-}$, cyan) or no P (orange) was compared to phosphate ($\text{PO}_4^{3-}$, dark blue) and is presented as the relative fold change (RFC). Significant differences (one-way ANOVA, $P < 0.05$) are indicated (distinct letter groups: x and y). Error bars represent standard deviation of the mean from three biological replicates.

Table 3C: The expression of phosphite ($\text{ptx}$) and phosphonate ($\text{phn}$)-metabolism genes in *Trichodesmium* transcriptomes.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>$\text{ptx}$ gene prevalence</th>
<th>$\text{phn}$ gene prevalence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary transcriptome</td>
<td>$\text{ptxABCD}$</td>
<td>$\text{phnDCEEGHIJKLM}$</td>
<td>(Pfreundt et al. 2014)</td>
</tr>
<tr>
<td>Gene expression profiles</td>
<td>$\text{ptxABCD}$</td>
<td>$\text{phnDCEEGHIJKLM}$</td>
<td>unpublished dataset</td>
</tr>
<tr>
<td>from high CO$_2$-adapted</td>
<td></td>
<td></td>
<td>NCBI accession: PRJNA237745</td>
</tr>
<tr>
<td><em>Trichodesmium</em></td>
<td></td>
<td></td>
<td>Chapter 5</td>
</tr>
<tr>
<td>Transcriptomic response of</td>
<td>$\text{ptxABCD}$</td>
<td>$\text{phnDCEEGHIJKLM}$</td>
<td></td>
</tr>
<tr>
<td><em>Trichodesmium</em> to different</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fe regimes</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In other bacteria the regulation of \textit{phn} (Wackett et al. 1987) and \textit{ptx} (White & Metcalf 2004) gene expression was demonstrated to be under the control of the Pho regulon (Santos-Beneit 2015). This is coordinated by a two-component mechanism involving the SphRS (referred to as PhoBR in \textit{E.coli}) transduction system which orchestrates the organisms’ response to low PO$_4^{3-}$ conditions (Tommassen et al. 1982; Juntarajumnong et al. 2007). This system comprises a histidine kinase sensor on the cell membrane (SphS) and a transcriptional response regulator (SphR).

\textit{Trichodesmium} is lacking the SphS component but it is suggested that SphR regulates the expression of P response operons. A putative binding site (PHO box) was identified upstream of P utilisation genes \textit{pstS}, \textit{sphX}, \textit{phoX} and \textit{phnD} (Su et al. 2007) while \textit{phoA} is 449 bp downstream of a sequence resembling that (Orchard et al. 2009) (Table 3D). This sequence is lacking the 3 base linker sequence between conserved domains and the authors concluded that this is explained by possible heterogeneity in the \textit{Trichodesmium} PHO box. To add to the observations by Orchard et al., (2009), a relevant short sequence directly upstream of \textit{ptxA} is also missing the 3 bp linker. Whether this justifies SphR regulation of \textit{ptxABCD} remains to be examined.

It is interesting to consider whether the separation of \textit{phnD} and \textit{ptxA}, by 862 and 449 bp respectively, from the rest of the genes in their respective ‘operons’ infers their distinct regulation to the rest of the cluster and/or whether more than one regulator is involved in their expression. In support of this, expression of \textit{ptxBCD} and \textit{phnD} seems to be independent to that of the remaining components in the unpublished transcriptome presented in Chapter 5 (Fig. 3.4). Also, anti-sense RNA appears to be expressed at the \textit{ptxA} locus adding another potential level of control to the expression of this gene (Fig. 3.4).
**Figure 3.4: Expression of genes along the ptx and phn clusters.** Genes identified from transcriptomic analysis (see Chapter 5) and visualised on ARTEMIS genome browser (Rutherford et al. 2000) are aligned along the forward DNA strand (5’ to 3’). Transcription is not uniform along the clusters with expression for the first gene of each being unique to the remaining genes. For the ptx cluster, detection of ptxA (Tery_0365) transcripts is minimal while ptxBCD (Tery_0366-0368) expression is more distinct. The opposite is true for the phn cluster which is overall expressed at low levels compared to phnD (Tery_4993), the first gene of the cluster. Anti-sense RNA appears to be transcribed at the ptxA locus.
Table 3D: Putative Pho box promoter binding sequences associated with P-metabolism involved genes.

<table>
<thead>
<tr>
<th>Teryno</th>
<th>Gene</th>
<th>Putative function</th>
<th>Predicted PHO box/PHO box resembling sequence</th>
<th>Position</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>3534</td>
<td>sphX</td>
<td>Phosphate binding</td>
<td>TTTGATAATtttTTTAACCTtgtt CTTAATCT</td>
<td>-164</td>
<td>Su et al. 2007</td>
</tr>
<tr>
<td>3537</td>
<td>pstS</td>
<td></td>
<td>TTTGATATtttTTTAACCTtgtt CTTAATCT</td>
<td>-2862</td>
<td>Su et al. 2007</td>
</tr>
<tr>
<td>3467</td>
<td>phoA</td>
<td>Alkaline phosphatase</td>
<td>TTTAACT TTTAACT TTTAACT</td>
<td>-449</td>
<td>Orchard et al. 2009</td>
</tr>
<tr>
<td>3845</td>
<td>phoX1</td>
<td></td>
<td>TTTAACCAaagCTTAAACGa aaTTTAAACC GAAAACAAGttTTTAACCTtttCTTAAACT</td>
<td>-173</td>
<td>Su et al. 2007</td>
</tr>
<tr>
<td>4993</td>
<td>phnD</td>
<td>Phosphonate binding</td>
<td>TTTAACCA TTTAACC</td>
<td>-435</td>
<td>Su et al. 2007</td>
</tr>
<tr>
<td>0365</td>
<td>ptxA</td>
<td>ATPase of phosphite transporter</td>
<td>TTTAACCA TTTAACC</td>
<td>-25</td>
<td>This study</td>
</tr>
</tbody>
</table>

3.2.5 Heterologous expression of *Trichodesmium ptxABCD* in *Synechocystis*.

As a stimulation of expression by PO$_4^{3-}$ stress is not sufficient to support involvement of the ptx genes in PO$_3^{3-}$ utilisation, a more direct approach was employed for their characterisation. Genetic manipulation techniques are not currently available for *Trichodesmium*, therefore to describe their function the ptxABCD genes were cloned into the pFLAG vector, which integrates inserted DNA in the *Synechocystis* genome so that expression is driven from the *psbA2* promoter. Homozygous strains in which all four genes (ptxABCD$^+$), only the ABC-transporter genes (ptxABC$^+$), or only the phosphite dehydrogenase gene (ptxD$^+$) replaced *psbA2* were obtained (Fig 3.5A). The sequence of integrated DNA was verified (Eurofins MWG Operon, Ebersberg, Germany) following amplification from the transformant genome and the expected pattern of gene expression for each strain was observed by end-point reverse transcription (RT)-PCR (Fig. 3.5B).

There are no homologues of the ptx genes in *Synechocystis* PCC 6803 (Kaneko et al. 1996), and in contrast to *Trichodesmium*, PO$_3^{3-}$ supplied as the sole source of P did not enhance the growth of the wild-type (WT) organism above that of a no-added P control (Fig. 3.6).
Figure 3.5: Confirmation of strain generation. (A) Amplification from *Synechocystis* WT and ptxABCD+, ptxABC+, ptxD+ stains, using primers upstream and downstream of the insertion to confirm full segregation at the *psbA2* locus. In each case the correct-sized product was obtained, and no WT band was present in the modified strains. HL1 = Hyperladder™ 1 kb (Bioline Reagents Limited, London, UK). (B) Expression of the *ptx* genes at the RNA level was confirmed by end-point RT-PCR. For each strain, the *ptx* genes were expressed or absent as expected. The *rnpB* gene was used as a reference gene and a no reverse transcriptase (RT) control confirmed RNA was not contaminated with gDNA.

Growth of the *Synechocystis* recombinant strains and the isogenic WT on PO$_3^-$ (175 μM Na$_2$HPO$_3$.5H$_2$O) or no P was compared to that on the standard PO$_4^-$ (175 μM NaH$_2$PO$_4$.H$_2$O) containing BG11 media under phototrophic conditions (no glucose). Highest growth rates were observed on PO$_4^-$ (and were used to correct for the relative growth rate of each strain in the other treatments) (Fig. 3.6A). All strains grew poorly when P was omitted, with residual growth under this condition likely due to PO$_4^-$ carryover from the starter culture. Conversely, when PO$_3^-$ was provided as the sole added P source, growth of the ptxABCD+ strain expressing both the ABC-transporter and the dehydrogenase grew to a comparable extent as when supplied with PO$_4^-$ (Fig. 3.6B). However, growth of ptxABC+ and ptxD+ on PO$_3^-$ remained low (Fig. 3.6B), indicating that expression of the ABC-transporter or dehydrogenase alone is not sufficient and both components are required to utilize this form of P.

To our knowledge, this is the first simultaneous demonstration of the functionality of both the transporter and dehydrogenase modules of a PO$_3^-$ utilization system from a photosynthetic organism, and clearly demonstrates the function of PtxABCD in *Trichodesmium*. 
Figure 3.6: Growth of *Synechocystis* with phosphite requires all four *ptx* genes. (A) Photoautotrophic growth of strains *ptxABCD*+, *ptxABC*+ and *ptxD*+ and the isogenic parent (WT) on phosphate (PO$_4$^3- ) was compared for each strain to growth on phosphite (PO$_3$^3- ) and P-deplete BG11 media (relative growth rate). Error bars represent the standard deviation from the mean of three biological replicates. Asterisks (***) indicate significant differences (Student’s t-test; P < 0.01) within strains. (B) Optical assessment of strains on the final day of the experiment after dilution 1:2.
Growth of the four strains in standard BG11 media with $\text{PO}_4^{3-}$ was uneven, with the $\text{ptxABCD}^+$ and $\text{ptxABC}^+$ growing at a faster rate than the WT and $\text{ptxD}^+$ ($P < 0.05$) (Fig. 3.7). These results indicate that P uptake by $\text{Synechocystis}$ cells is not functioning at maximal efficiency and is therefore a limiting step during exponential growth. Therefore, the presence of a functional transporter expressed at high levels (expression controlled by $\text{psbA2}$ promoter) allows faster growth rates.

Figure 3.7: Strains of $\text{Synechocystis}$ grow unevenly on phosphate. Growth rates of $\text{ptxABCD}^+$, $\text{ptxABC}^+$ and $\text{ptxD}^+$ and the isogenic WT parent were compared during growth on standard, phosphate ($\text{PO}_4^{3-}$) containing BG11. Error bars represent the standard deviation from the mean of three biological replicates. Letter groups (x and y) are significantly different (Student’s t-test, $P < 0.05$) to each other.

However, these observations can only be valid if (a) some $\text{PO}_3^{3-}$ is present in the media, either through contamination or reduction of $\text{PO}_4^{3-}$, and this is taken up by the $\text{ptxABCD}$ transporter, and/or (b) the transporter can also take up $\text{PO}_4^{3-}$ with lower specificity/efficiency. For the former alone to explain the observations, the cell must prepossess enzymes that can carry out oxidation of $\text{PO}_3^{3-}$ to $\text{PO}_4^{3-}$ as $\text{ptxABCD}^+$ carries the ABC transporter genes but lacks the putative phosphite dehydrogenase ($\text{ptxD}$).
As seen previously (Fig. 3.6), $\text{PO}_3^{3-}$ cannot be efficiently used by $\text{ptxABCD}^+$ as a source of P therefore utilisation of $\text{PO}_3^{3-}$ is not a possible explanation. Consequently, $\text{ptxB}$ is hypothesised to also bind $\text{PO}_4^{3-}$ and this is currently under investigation (Bisson et al., manuscript in preparation). Although the higher growth rate of $\text{ptxABCD}^+$ compared to $\text{ptxABCD}^+$ is not statistically significant, the possibility of uptake of trace $\text{PO}_3^{3-}$ concentrations, present in the media, by both and its subsequent utilisation by $\text{ptxABCD}^+$ alone, cannot be excluded.

### 3.2.4 Environmental prevalence of $\text{ptx}$ genes in *Trichodesmium*

To identify the environmental prevalence of $\text{ptxABCD}$, nucleotide sequences of the genes from *T. erythraeum* IMS101 were compared against all available *Trichodesmium* metagenomic and metatranscriptomic datasets (Table 3E).

A contig sequence (TCCM\_contig00441) was identified in a *Trichodesmium* community metagenome assembly from the Atlantic Ocean (Walworth et al. 2015) that possessed putative open reading frames encoding homologues of $\text{ptxA}$ (bit score: 401, identity: 86%), $\text{ptxB}$ (bit score: 316, identity: 76%) and $\text{ptxC}$ (bit score: 420, identity: 93%).

Furthermore, metatranscriptomic reads aligning to all four $\text{ptx}$ genes (percentage identity > 90%) could be identified in samples from two sites in a dataset obtained from the Amazon River Plume (Hilton et al. 2015). Similarly, $\text{ptxAB}$ were identified in metatranscriptome analysis from the New Caledonian Lagoon in the South West Pacific (as indicated in Spungin et al. 2016). The raw data is not yet available online to search for the presence of $\text{ptxD}$. Reads matching the $\text{ptx}$ genes were not identified in a 454 metatranscriptomic dataset from the South Pacific Ocean (Hewson et al. 2009); however, the overall coverage of *Trichodesmium* coding sequences in this study was very low.

Identification of prevalence of $\text{ptxABCD}$ genes in *in situ* metagenomic and metatranscriptomic datasets confirms that they are both present and expressed in diverse *Trichodesmium* species in the field, further supporting the role of $\text{PO}_3^{3-}$ uptake in defining *Trichodesmium*’s ecological niche.
Chapter 3: Phosphite utilisation by *Trichodesmium*

### Table 3E: Environmental prevalence of ptx genes in *Trichodesmium*

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Location</th>
<th>Identified</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>In situ</em> metagenome</td>
<td>Atlantic Ocean</td>
<td>ptxABC</td>
<td>Walworth et al. 2015</td>
</tr>
<tr>
<td><em>In situ</em> metatranscriptomes</td>
<td>Amazon River Plume</td>
<td>ptxABCD</td>
<td>Hilton et al. 2015</td>
</tr>
<tr>
<td></td>
<td>South Pacific Subtropical Gyre</td>
<td>No</td>
<td>Hewson et al. 2009</td>
</tr>
<tr>
<td></td>
<td>New Caledonia Lagoon</td>
<td>ptxABC*</td>
<td>Spungin et al. 2016</td>
</tr>
</tbody>
</table>

* The presence of ptxD is not indicated in the paper and the raw data is not currently available for manual identification

#### 3.3 Conclusions and further discussion

Following evidence supporting the involvement of marine organisms including *Trichodesmium* in a dynamic P redox cycle (Dyhrman et al. 2006; Martínez et al. 2012; Van Mooy et al. 2015), there is emerging interest in understanding the mechanisms involved in the biological processes responsible, and the extent that biology influences or is influenced by oceanic processes. Work developed in the context of this Chapter was aimed at deciphering whether *Trichodesmium*’s physiology can support growth on reduced inorganic P and characterising the molecular mechanism underlying this.

It was established that the organism can grow on PO$_3^{3-}$ as its sole source of P, potentially extending the oceanic locations currently considered as *Trichodesmium*’s ecological niche. This possibility cannot be presently tested since no measurements of PO$_3^{3-}$ concentrations in the ocean have been reported. The development of suitable analytic methods for the quantification of reduced P sources in the marine environment is necessary to establish its significance for the organism’s lifestyle.

Using heterologous expression in *Synechocystis* PCC 6803, the putative *Trichodesmium* ptxABCD operon, which was found to be upregulated under PO$_4^{3-}$-limitation, is identified as the molecular system responsible for accessing PO$_3^{3-}$ with both the transporter (ptxABC) and phosphite dehydrogenase (ptxD) required for the process. Our results highlight the power of heterologous expression for studying the function of *Trichodesmium* proteins in genetically tractable cyanobacterial species, as previously shown by HetR complementation in *Anabaena* (Kim et al. 2011).
By analogy to the heterologous expression method, it is interesting to speculate whether virally mediated lateral gene transfer, which is common in microbial communities (Lindell et al. 2005; Zeidner et al. 2005), promotes natural acquisition of \( \text{ptx} \) genes in the contemporary ocean, and consequently question how abundant these genes are across different species/ecotypes. The genes are identified to be present and expressed \textit{in situ} both in \textit{Trichodesmium} (this study) and \textit{Prochlorococcus} (Martínez et al. 2012).

The ability to catabolise \( \text{PO}_3^{3-} \), as demonstrated here, is also connected to the organism’s ability to synthesize it (Van Mooy et al. 2015) as there would be little or no use of production of an inaccessible compound. The question of the purpose of \( \text{PO}_3^{3-} \) production remains, but there have been suggestions of possible use as an antibiotic by organisms (Karl 2014) due to its fungicide qualities (Griffith et al. 1989). In addition, if the ability to use \( \text{PO}_3^{3-} \) has, contrary to the current surge of evidence, not spread amongst marine organisms, production of \( \text{PO}_3^{3-} \) could increase the likelihood that after its release it would be reused by the same species.

To conclude, results presented in this chapter support that the \( \text{PO}_3^{3-} \) utilising ability of \textit{Trichodesmium} is important to determining the organism’s success in its natural environment, with \textit{Trichodesmium} itself potentially a major contributor to a dynamic biological P redox cycle.

As a final note, it is interesting to discuss the potential biotechnological applications of this work. A by-product of the research conducted for purposes of this project, the \( \text{PO}_3^{3-} \) utilising \textit{Synechocystis} PCC 6803 stain \( \text{ptxABCD}^{+} \), was observed to be particularly efficient in the use of \( \text{PO}_3^{3-} \) as its only source of P. Algae and cyanobacteria are currently beginning to be used for wastewater treatment in combination with biofuel production as the need for green energy increases (Christenson & Sims 2011; Nozzi et al. 2013; Lynch et al. 2015) with synthetic biology employed to optimise the productivity and viability of these methods (Georgianna & Mayfield 2012). Since phosphites are used as fungicides, pesticides and herbicides as well as fertilisers and could be identified in wastewater treatment water (Yu et al. 2015) the use of genetically engineered \( \text{PO}_3^{3-} \) utilising strains for wastewater treatment could be an efficient way of dealing with residual P potentially contributing to eutrophication of riverine/coastal water. In addition, the potential production of NADH during oxidation of \( \text{PO}_3^{3-} \) by the \( \text{ptxD} \) could drive NAD(P)H with possible applications in enhancing the activity of biophotovoltaic devices (Anderson et al. 2016).
Chapter 4

The function of Tery_3377 (IdiA)
Iron transport or PSII protection?

4.1 Introduction

In addition to phosphorus (P) limitation, oceanic primary production is challenged by an insufficiency of the micronutrient iron (Fe) (Moore et al. 2013), which is necessary for construction of the cell’s photosynthetic machinery (Geider & La Roche 1994). Due to its additional use as a component of the nitrogenase enzyme (Sañudo-Wilhelmy et al. 2001; Shi et al. 2007; Richier et al. 2012), Fe is essential for the prevalence of diazotrophy in the ocean (Mills et al. 2004; Moore et al. 2009; Chappell et al. 2012). Therefore, organisms have developed diverse strategies that permit survival at fluctuating or chronically depleted Fe concentrations. These include intricate Fe uptake mechanisms and protective responses against elevated production of reactive oxygen species (ROS) generated as a result of Fe-stress (Peers & Price 2004). Iron deficiency induced protein A (IdiA) homologues have been suggested to participate in both of these processes.

Proteins with similarity to IdiA are found in organisms of all domains of life from archaea (eg. *Natrialba chahannaoensis*) to eukaryotes (e.g. the polychaete *Capitella teleta* and the Tibetan antelope *Pantholops hodgsonii*), but has only been studied in a few pathogenic bacteria and cyanobacteria species.
The first homologue of IdiA, referred to as FbpA (ferric ion binding protein A), was discovered in pathogenic bacteria and was characterised as a periplasmic Fe\(^{3+}\)-binding protein, which in conjunction with transmembrane (FbpB) and ATPase (FbpC) components, forms an ATP binding cassette (ABC) transporter for transport of ferric iron (Fe\(^{3+}\)) (Krewulak & Vogel 2008). In cyanobacteria, an IdiA homologue was first identified in *Synechococcus elongatus* PCC 6301/PCC 7942 and was unexpectedly shown to be intracellular. It was associated with the thylakoid membranes and has a role in protection of photosystem II (PSII) under oxidative stress caused by Fe-deficiency (Michel et al. 1996; 1998; Exss-Sonne et al. 2000). The marked upregulation of IdiA homologues observed in several marine organisms, including *Trichodesmium*, under Fe-stress, lead to the consideration and use of IdiA as an *in vivo* marker of Fe-limitation (Rivers et al. 2009; Chappell & Webb 2010; Saito et al. 2014; Snow et al. 2015; Walworth et al. 2016).

In *Synechocystis* sp. PCC 6803 (hereafter *Synechocystis*) there are two paralogues of the protein referred to in literature as FutA1 (Slr1295) and FutA2 (Slr0513). The *futA1* and *futA2* genes as well as the permease (*futB*, *slr0327*) and the ATPase (*futC*, *slr1878*) homologues are all separately localised relative to each other in the *Synechocystis* genome. Deletion of the *futB* and *futC* components compromise the organisms growth, while the Δ*futA1*futA2 double mutant is severely Fe-limited compared to the WT (5% Fe uptake compared to the WT (Katoh et al. 2001a). Deletions of *futA1* or *futA2* independently resulted in reduced Fe uptake (37% and 84% residual activity compared to WT respectively) and impaired growth under Fe-deficiency (Katoh et al. 2001b). Contradicting the reported decrease of Fe uptake in Δ*futA1* by this study, Kranzler et al. (2014) suggest an increase compared to the WT (Table 4A).

**Table 4A:** Phenotypes previously reported in *Synechocystis* Δ*futA1* and Δ*futA2* mutants

<table>
<thead>
<tr>
<th>Δ<em>futA1</em></th>
<th>Δ<em>futA2</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Low PSII activity (Tolle et al. 2002)</td>
<td>Plastocyanin- Cytochrome c(_{553}) (PC) switch not performed (Waldron et al. 2007)</td>
</tr>
<tr>
<td>Earlier progression to PSI-IsiA (Tolle et al. 2002)</td>
<td>Lower cytochrome c oxidase activity (Waldron et al. 2007)</td>
</tr>
<tr>
<td>Fe uptake affected (reduced in Katoh et al. 2001b, increased in Kranzler et al. 2013)</td>
<td>Decreased Fe uptake (Katoh et al. 2001b, Kranzler et al. 2013)</td>
</tr>
<tr>
<td>Lower Fe(^{3+})-FutA2 in periplasm (Badarau et al. 2007)</td>
<td>Pigmented supernatant (Waldron et al. 2007)</td>
</tr>
</tbody>
</table>
Despite one report suggesting that FutA1 and FutA2 bind Fe$^{2+}$ (Koropatkin et al. 2007), both have been shown to bind Fe$^{3+}$ (Katoh et al. 2001b; Waldron et al. 2007; Badarau et al. 2008). Therefore, both were initially predicted to function as semi-redundant periplasmic Fe-binding components involved in Fe$^{3+}$ uptake in conjunction with the FutB and FutC subunits (Katoh et al. 2001b). However, further research suggested that only FutA2 is found in the periplasm, with FutA1 localised intracellularly (Fig. 4.1) (Fulda et al. 2000; Tölle et al. 2002; Waldron et al. 2007).

![Figure 4.1: Current model of the localisation of *Synechocystis* Fut paralogues. FutA1 (left) is thought to be intracellular, attached to PSII at the thylakoid membranes, and have protective properties under oxidative stress caused by Fe-deficiency. FutA2 (right) is periplasmic and is postulated to act as the Fe$^{3+}$ binding protein of the Fut ABC-transporter for Fe$^{3+}$ uptake across the plasma membrane.](image)

Functionally, FutA1 has been proposed to protect PSII under Fe-deficiency (Tölle et al. 2002), similarly to IdiA from *Synechococcus*. Indeed, PSII was more susceptible to inactivation in the ΔfutA1 mutant, which had a reduction in photosynthetic pigment content and oxygen evolution, and showed more advanced attachment of iron-stress-induced protein A (IsiA) (Fe-deficiency induced antenna) to photosystem I (PSI) under mild Fe-limitation (Table 4A) (Tölle et al. 2002). Conversely, the periplasmic FutA2 potentially interacts with FutBC to form an active Fe$^{3+}$ transporter, and appears to have a role in preventing aberrant associations of Fe with binding sites of other metal transporters (Waldron et al. 2007). In particular, cytoplasmic copper containing enzymes are affected in the ΔfutA2 mutant, which has lower cytochrome c oxidase activity and uses cytochrome c$_6$ under conditions where the WT uses the copper-containing plastocyanin (Table 4A) (Durán et al. 2004; Waldron et al. 2007).
Unlike *Synechocystis*, *Trichodesmium erythraeum* IMS101 (hereafter *Trichodesmium*) has only one *futA* homologue in its genome, encoded by *Tery_3377* (Chappell & Webb 2010). This is the protein identified in a search for IdiA homologues by Michel & Pistorius (2004), although the authors wrongly reported it as being Tery_0244 (annotated as a hypothetical protein). Increased expression of Tery_3377 under Fe-stress has been reported at both the transcriptional (see Chapter 5 and Shi et al. 2007; Chappell & Webb 2010), and protein (Webb et al. 2001; Snow et al. 2015) levels.

This study attempts to address the question of the function of *Trichodesmium* Tery_3377 relative to the two *Synechocystis* FutA paralogues. Bioinformatic and phylogenetic analysis is first utilised to explore the similarity of the *Trichodesmium* protein to the *Synechocystis* and other bacterial homologues. Subsequently, using molecular approaches the cellular localisation and function of the protein are examined. For this, *Synechocystis* is used as a vehicle for the expression of the *Trichodesmium* protein or its signal sequence fused to green fluorescent protein (GFP).

### 4.2 Results and discussion

#### 4.2.1 Bioinformatic analysis of Tery_3377

Although largely assumed to be involved in Fe transport and variably referred to as FutA or IdiA, no data is available regarding the *in vivo* function of Tery_3377. Similarly to the *Synechocystis* futA genes, Tery_3377 is localised at an independent locus to the ABC transporter transmembrane (*Tery_3223*) and ATPase (*Tery_3222*) components in the *Trichodesmium* genome. The Tery_3377 protein has 57% identity to FutA1 and FutA2 (Table 4B).

To understand how homologues of the protein are grouped based on their sequence, maximum likelihood phylogenetic analysis was performed to include the *Trichodesmium* and the two *Synechocystis* variants along with their closest homologue (from *Lyngbya* sp. PCC 8106, *Phormidium willei* and *Cyanothece* sp. CCY 0110 for the *Trichodesmium* Tery_3377 and *Synechocystis* FutA2 and FutA1 respectively) as identified using BlastP (Altschul et al. 1990). Homologues from *Synechococcus elongatus* PCC 6301 and PCC 7942 as well as *Prochlorococcus marinus* MIT 9301 and *Anabaena* sp. PCC 7120 were also included in the analysis (Table 4B, Fig. 4.2).
Table 4B: IdiA/FutA/Fbp homologues of selected bacteria. Putative signal sequences and the first 4 amino acid residues of the mature protein as predicted from Signal P v4.1 (v3.0 sensitivity) and TatP v1.0 servers (Bendtsen et al. 2005; Petersen et al. 2011). Where relevant an alternative signal sequence end is marked (*). The likelihood that these are functional signal sequences, as assessed in each server, is also reported. Twin arginine (RR) residues, other Tat conserved amino acids, positively charged residues and putative lipobox sequence are indicated. The localisation of the protein is stated if previously studied (Michel et al. 1998; Fulda et al. 2000; Tölle et al. 2002; Waldron et al. 2007; Khambati et al. 2010). The amino acid (aa) identity of each protein to Tery_3377 was identified through BlastP (Altschul et al. 1990).

<table>
<thead>
<tr>
<th>Organism</th>
<th>Protein</th>
<th>Predicted Signal Sequence</th>
<th>SignalP v4.1 (v3.0 sensitivity)</th>
<th>TatP v1.0</th>
<th>Putative signal sequence length (aa) (suggested alternative length)</th>
<th>Reported location</th>
<th>Identity to Tery_3377 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Trichodesmium erythraeum IMS 101</em></td>
<td>Tery_3377</td>
<td>MTITRRVFGLGTGTAIA*ELGKSNRGLAQSGA</td>
<td>no</td>
<td>no</td>
<td>31 (21)</td>
<td>unknown</td>
<td>100</td>
</tr>
<tr>
<td>Synecocystis sp. PCC 6803</td>
<td>Str0513 (FutA2)</td>
<td>MTTIKISRTFVGGTALTAVANLPRRBASAQRSRT</td>
<td>yes</td>
<td>yes</td>
<td>31</td>
<td>periplasm</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>Stry1295 (FutA1)</td>
<td>MVQKLSRLFLSISIGTAFTTVGSGQLLSCGQSPDAPIADTPG</td>
<td>yes</td>
<td>no</td>
<td>37</td>
<td>intracellular</td>
<td>57</td>
</tr>
<tr>
<td><em>Synechococcus elongatus PCC 6301</em></td>
<td>Syc1920_d</td>
<td>MHHNLSPLSMSESMSRRDFLLGGTALATLLDSFGRWRBAEAAGGE</td>
<td>no</td>
<td>no</td>
<td>45</td>
<td>intracellular</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>Syc0146_c</td>
<td>MSLHRWSTGLAALLSTGGLAC*CSGDKV</td>
<td>yes</td>
<td>no</td>
<td>25 (22)</td>
<td>unknown</td>
<td>53</td>
</tr>
<tr>
<td><em>Prochlorococcus marinus MIT 9301</em></td>
<td>P9301_13611</td>
<td>MQKIKKIIIFSFTCTFELLNVNIPANSTIKE</td>
<td>yes</td>
<td>no</td>
<td>26</td>
<td>unknown</td>
<td>36</td>
</tr>
<tr>
<td>Anabaena sp. PCC 7120</td>
<td>Alg1382</td>
<td>MVSHQRQISSAITAIVLGLGLGIAVAAATPK</td>
<td>no</td>
<td>no</td>
<td>27</td>
<td>unknown</td>
<td>28</td>
</tr>
<tr>
<td>Haemophilus influenza</td>
<td>Hl_0097</td>
<td>MQFKHFKLALAAALAFSANSFADI TV</td>
<td>yes</td>
<td>no</td>
<td>23</td>
<td>periplasm</td>
<td>29</td>
</tr>
</tbody>
</table>
Figure 4.2: Maximum likelihood phylogenetic analysis of Tery_3377 homologues from selected bacteria. *Synechocystis* FutA1 and FutA2 proteins and the Trichodesmium Tery_3377 (blue) are compared. Their closest homologues identified using BlaspP (Altschul et al. 1990) as well as those from *Synechococcus elongatus* PCC 6301/7942, *Prochlorococcus marinus* MIT 9301, *Anabaena* PCC 7120 and pathogenic bacteria *Haemophilus influenzae* and *Neisseria meningitides* are included in the analysis. The graphical representation was produced using TreeDyn v198.3 (Chevenet et al. 2006) and branch probabilities reported are calculated using an approximate likelihood ratio test (aLRT, Anisimova & Gascuel 2006). The proteins form two groups: putative Fe transporters (red), and those with possible intracellular function (green).
Despite early studies only reporting the occurrence of a single homologue in the *Synechococcus elongatus* PCC 6301/PCC 7942 genome (Michel et al. 1996; Michel et al. 1998; Michel et al. 1999; Exss-Sonne et al. 2000; Michel & Pistorius 2004) the presence of two independent homologues was later revealed (Rivers et al. 2009; Chauvat & Cassier-Chauvat 2012). However, comparative studies for the two paralogues were not performed. The phylogenetic analysis completed here identifies that the *Synechococcus* proteins cluster closer to their corresponding *Synechocystis* homologues (FutA1 or FutA2) than to each other (Fig. 4.2). Genes encoding for the homologues that clustered with *Synechocystis* FutA2 (annotated as ‘iron transport system substrate-binding protein’ and ‘futA2’ for *S. elongatus* sp. PCC 6301 and 7942 respectively) are colocalised with other members of an ABC transporter in the *Synechococcus* genome. This is contrasting what is observed for both *Synechocystis* futA homologues and *Trichodesmium Tery_3377*, which are not colocalised with the putative partner transporter genes in their respective genomes.

The second protein homologue of the two species, annotated as ‘IdiA’ and previously identified to have an intracellular role in PSII protection (Michel et al., 1996, 1998; Michel and Pistorius 2004), seems to branch out of the Fe transporter group (Fig. 4.2), as does the *Synechocystis* FutA1 protein. Interestingly, the *Trichodesmium Tery_3377* regarded as an Fe transporter is more closely related to FutA1 and IdiA compared to the Fe transporter proteins of *Synechocystis* and *Synechococcus*.

Similarly to *Trichodesmium*, and unlike *Synechococcus, Prochlorococcus* seems to only possess a single copy of the protein, the gene of which is not colocalised with other transporter components but is annotated and regarded (Thompson et al. 2011) as a putative Fe transporter. The phylogenetic analysis performed here indicates a closer relationship to FutA2 and other transporter homologues (Fig. 4.2) compared to FutA1 and thus supports its functional annotation. Previously identified high sequence divergence of this *Prochlorococcus* homologue and its localisation next to tRNA sequence, are indicative of potential subjection to horizontal gene transfer and loss events (Thompson et al. 2011).

Interestingly, the identified homologue of the model freshwater diazotroph *Anabaena* sp. PCC 7120 appears to be more related to the pathogenic bacteria *Hemophilus influenzae* and *Neisseria meningitides* FbpA than the other cyanobacterial homologues included in the analysis. In addition, similarly to the pathogenic bacteria variant and the *Synechococcus futA2*, it is localised upstream of the ABC transporter permease and ATPase homologues, making its assignment as a transporter unambiguous.
By inspecting the predicted protein signal sequences, we can potentially draw inferences to where proteins will be directed in the cell. Export of proteins to the periplasmic space can be achieved either through the general secretory (Sec) pathway, whereby proteins fold after their transport across the inner membrane, or the twin arginine translocase (TAT) system, which transports proteins previously folded in the cytoplasm (Natale et al. 2008). The latter is distinguished through the conserved N-terminal motif SRRXFLK in which the two consecutive arginine residues (giving the TAT system its name) are almost always present (Lee et al. 2006). The *Synechocystis* FutA1 and FutA2 homologues as well as *Trichodesmium* Tery_3377 possess amino acids with similarity to this motif, and all include the conserved arginine residues (Table 4B). In addition, TAT substrate proteins commonly express AXA at the cleavage site and a positive charge of +2 or higher at the C-terminal region is thought to promote TAT specificity (Cristóbal et al. 1999; Blaudeck et al. 2003; Lee et al. 2006). This signal appears to be present for the *Synechocystis* FutA2 but not FutA1 homologue while Tery_3377 expresses AXA but no positively charged residues (Table 4B).

Signal sequence prediction server SignalP 4.1 (sensitivity of version 3.0, Petersen et al. 2011) does not positively identify a signal sequence for Tery_3377 unlike for the two *Synechocystis* FutA proteins. TAT-signal prediction server TatP 1.0 (Bendtsen et al. 2005) designed to identify proteins exported through the TAT system gives a positive result with the *Synechocystis* FutA2 protein but not FutA1 or the *Trichodesmium* Tery_3377 (Table 4B). For Tery_3377, this could be explained by either an intracellular role in PSII protection under Fe-limitation analogous to *Synechocystis* FutA1 and *Synechococcus* IdiA homologues or a divergent signal sequence for export to the periplasm. Despite both of the *Synechocystis* FutA homologues containing the twin arginine residues, the observed dissimilarity in localisation (Fulda et al. 2000; Tölle et al. 2002), also predicted by the TatP server, could be due to loss of the TAT signal sequence C-terminal peptidase recognition site required for cleavage in FutA1 (Table 4B).

The positive prediction of a Tat signal sequence for FutA2 is unique amongst the bacteria considered here and in addition to *Synechocystis* FutA1 and *Trichodesmium* Tery_3377, no other predicted signal sequences contain two arginine amino acids in the N-terminal region. Therefore, the periplasmic Fe-binding *Haemophilus influenza* FbpA (Adhikari et al. 1995; Kirby et al. 1997; Anderson et al. 2004) and potentially the *Prochlorococcus marinus* str. MIT 9301, P9301_13611, *Anabaena* PCC 7120, Alr1382, and *Synechococcus elongatus* PCC 6301 transporter homologue, Syc0146, are likely to be translocated to the periplasmic
space via the Sec pathway. Although the *Prochlorococcus marinus* str. MIT 9301 and *Synechococcus elongatus* PCC 6301 homologues, P9301_13611 and Syc0146 respectively, are positively predicted to possess a signal sequence by SignalP 4.1 (sensitivity of version 3.0 Petersen et al. 2011), Alr1382 of *Anabaena* PCC 7120 is not (Table 4B) contradicting its similarity to periplasmic FutA2-like homologues (Fig. 4.2) and colocalisation with FutBC-transporter genes in its genome (Stevanovic et al. 2012).

Bacterial lipoproteins are a large class of functionally diverse membrane associated proteins (Nakayama et al. 2012). They contain substrate binding proteins of ABC transporters but also the cyanobacterial PSII associated PsbP, PsbQ, and Psb27 (Nowaczyk et al. 2006; Fagerlund & Eaton-Rye 2011; Mizusawa & Wada 2012). Cleavage by type II (or lipoprotein) signal peptidase (Lsp) occurs at the C-terminal recognition site which commonly possess L(A/S/T)(G/A) as the last three amino acids and C always as the first residue of the mature protein (Hutchings et al. 2009). The *Synechococcus elongatus* PCC 6301 Syc0146 is the only homologue of the ones analysed which appears to possess a sequence region resembling the lipobox motif for Lsp recognition (but Lsp cleavage would result in a signal sequence shorter than the one predicted by the SignalP server) (Table 4B).

An overview: Phylogenetic analysis (Fig. 4.2) indicates that the cyanobacterial and pathogenic bacteria FutA/IdiA/FbpA homologues analysed can be grouped in FutA1-like and FutA2-like clusters. *Trichodesmium* Tery_3377 although considered an Fe transporter seems more similar to intracellular FutA1-like homologues indicating the possibility for an alternative function related to PSII protection under Fe-stress-induced oxidative stress. If this is the case, then another yet to be identified Fe$^{3+}$ uptake mechanism must be operating in this organism.

To investigate the possible cellular localisation of proteins the signal sequences were compared (Table 4B). Using online prediction software *Trichodesmium* Tery_3377 is not identified to possess a functional signal sequence. However, the same is true for the *Anabaena* PCC 7120 homologue which phylogenetic analysis suggests to be similar to periplasmic FutA2-like proteins. Therefore, bioinformatic analysis alone is not sufficient to resolve the question of the localisation of Tery_3377.

### 4.2.2 Tery_3377 cellular localisation

In order to provide direct visual evidence of Tery_3377 localisation and compare it to FutA1 and FutA2 in *Synechocystis*, constructs in which each of the possible (using TatP v1.0)
longest signal peptides (Table 4B) in addition to the predicted first 4 residues of the mature protein fused to GFP were generated using overlap extension (OLE)-PCR. Selection of the construct was achieved through the chloramphenicol acetyl transferase gene (cat<sup>b</sup>). The GFP-tagged DNA constructs were integrated at the psbA2 (sl1311) locus in Synechocystis such that expression was under the control of the psbA2 promoter. The E. coli TMAO reductase (TorA) signal sequence, previously shown to transport GFP into the periplasm in both E. coli (Thomas et al. 2001) and Synechocystis (Spence et al. 2003), was used as positive control for translocation across the cytoplasmic membrane by the TAT system. Strains expressing GFP fused signal sequences (ss-GFP) of the Trichodesmium Tery 3377, Synechocystis futA1 and futA2, and the E. coli control torA were successfully generated (Fig. 4.3).

Figure 4.3: The GFP tagged-signal sequence strains. (A) Genes for GFP and cat<sup>b</sup> were assembled next to signal sequences and the constructs replaced psbA2 in the Synechocystis genome. (B) Signal sequences as follows: Synechocystis futA1 (1) and futA2 (2), Trichodesmium Tery 3377 (3) and the E.coli torA (4), were amplified from each resulting strain with primers clone_F and clone_R and consequently sequenced to validate the inserted sequence. Control reactions of the WT (5) carried out with the same primers identified no product amplification as expected. Lane M = Hyperladder I (Bioline, London, UK).
Figure 4.4: Localization of signal sequence-GFP constructs in *Synechocystis*. Using confocal microscopy the WT (A) and strains with GFP fused to the signal sequence of the *E.coli* control TorA (B), *Synechocystis* FutA1 (C) and FutA2 (D) and *Trichodesmium* Tery_3377 (E) were imaged under identical settings through excitation at 488 nm where autofluorescence (represented by phycocyanin) emits at 600-700 nm (Ai-Ei) and GFP at 510-530 nm (Aii-Eii). The merged images for each strain (Aiii-Eiii) reveal periplasmic localisation of GFP by green fluorescent halos around the periphery of the cells in all strains (B-E) apart from the WT (A).
Confocal microscopy was used to image autofluorescence (em: 600-700 nm) and GFP fluorescence (em: 510-530 nm) simultaneously. Overlaying the two images revealed fluorescent halos in all strains except the WT, suggesting a periplasmic or cytoplasmic membrane localisation of GFP (Fig. 4.4). Our data contradict Tölle et al. (2002) that suggest the absence of any FutA1 in the periplasm. The intensity of halos differed between strains with FutA2ss-GFP, expected to be periplasmic (Fulda et al. 2000; Tölle et al. 2002; Waldron et al. 2007), being distinctly brighter than FutA1ss-GFP and closely resembling the TorAss-GFP control. The *Trichodesmium* Tery_3377ss-GFP is also prominently localised to the periplasm (Fig. 4.4). These observations can be explained as export of all ss-GFP fusions to the periplasm (possibly much less for FutA1) and provides the first direct evidence of the *Trichodesmium* Tery_3377 localisation as periplasmic, supporting a function as an Fe transporter.

### 4.2.3 Mode of Tery_3377 export to the periplasm

Using site-directed mutagenesis (SDM), the twin arginine residues of Tery_3377 possibly part of a TAT signal peptide, were substituted by lysine (K) (also a large, basic and positively charged amino acid) residues. In that way, a second Tery_3377 signal sequence fusion to GFP was constructed (Tery_3377ss-KK). The double arginine to lysine substitution is known to prohibit the TAT signal sequence recognition and consequent export to the periplasm (Buchanan et al. 2001; Ize et al. 2002).

The expression of the constructs at the protein level was identified using immunoblotting with anti-GFP primary antibody (Fig. 4.5). Product size differed between strains with Tery_3377ss-GFP being smaller than the Tery_3377ss-KK-GFP. This is in agreement with cleavage of the signal sequence from GFP, occurring during translocation, when the sequence is intact (Tery_3377ss-GFP). The Tery_3377ss-KK-GFP signal sequence is potentially rendered unrecognisable to the TAT system and thus remains attached to GFP and presumably intracellular. A lower signal of GFP expression in this treatment could be indicative of intracellular degradation of non-transported substrates.

Visualisation of the ss-GFP construct expression with epifluorescence microscopy revealed loss of the fluorescent halos around the cells’ periphery in the case of Tery_3377ss-KK-GFP (Fig. 4.6), compared to Tery_3377ss-GFP, confirming the immunoblot results and characterising Tery_3377 as a TAT system substrate.
Figure 4.5: Identification of GFP expression in *Synechocystis* strains. Immunoblot analysis using primary antibody anti-GFP on protein samples extracted from WT *Synechocystis* (1), strains containing GFP only (2), TorAss-GFP (3), Tery_3377ss-GFP (4) and Tery_3377ss-KK-GFP (5). Lane L = Ladder MagicMark™ XP (Thermo Fisher Scientific, Massachusetts, USA).

Figure 4.6: Localization of the *Trichodesmium* Tery_3377 signal sequence-GFP constructs in *Synechocystis*. Imaged using epifluorescence microscopy WT cells (A) and *E.coli* TorAss-GFP fusions (B) are presented as a control to *Trichodesmium* Tery_3377ss-GFP (C) and Tery_3377ss-KK-GFP (D). Tery_3377ss-GFP and Tery_3377ss-KK-GFP are also visualised together for direct comparisons (E).
4.2.4 *Synechocystis* futA mutants generation and physiology

To study the function of Tery_3377 by relating it to that of FutA1 and FutA2, *Synechocystis* futA1 and futA2 deletion mutants (ΔfutA1 and ΔfutA2) were generated. PCR confirmed their full segregation using a combination of flanking, gene-specific and antibiotic resistance cassette specific primers (Fig. 4.7). The ΔfutA1/futA2 double mutant was also constructed, and segregation was determined in a similar manner (data not shown).

![Diagram](image)

**Figure 4.7**: *Synechocystis* ΔfutA1 and ΔfutA2 deletion mutants were generated. (A) Representation of the futA1/futA2 locus arrangement in WT and deletion mutants of *Synechocystis*. The position of primers used to confirm complete segregation is indicated. (Bi) Screening of ΔfutA1 and comparison with the WT through PCR using primers futA1_ P1, P4, P5 and zeo_R confirmed the full segregation of the mutant. (Bii) Similarly, ΔfutA2 and WT *Synechocystis* were screened with primers futA2_ P1, P4, P5 and cat_R2 to confirm the full segregation of the mutant. Lane M = Hyperladder I (Bioline, London, UK).
The physiology of the mutants was assessed, and similar to previous reports (Katoh et al. 2001a) it was observed that growth rate was reduced in the ΔfutA1/ΔfutA2 double mutant by about 50% but not in the ΔfutA1 or ΔfutA2 compared to the WT (Fig. 4.8A). Also an Fv/Fm (estimate of the apparent PSII photochemical quantum efficiency) reduction in both the ΔfutA1/ΔfutA2 and the ΔfutA2 mutants (60 and 25% respectively compared to the WT), as opposed to a small increase for ΔfutA1, is reported here for the first time (Fig. 4.8B). The divergence of ΔfutA1 and ΔfutA2 Fv/Fm could be triggered by increased or decreased Fe uptake and reduction respectively compared to the WT, as previously reported by Kranzler et al. (2014).

Figure 4.8: Physiology of FutA mutant strains in *Synechocystis*. (A) The growth of ΔfutA1/ΔfutA2, measured over 11 days (OD750) was severely compromised compared to that of WT, ΔfutA1 and ΔfutA2. (B) The photosynthetic efficiency (Fv/Fm) of each strain differed to the others during exponential growth (day 4) (distinct letters mark significant differences between strains) (GLM and Tukey-test, CI=95%). Error bars represent standard deviations from the means of three biological replicates.
4.2.5 Expression of Tery_3377 in *Synechocystis* mutants

The codon optimised *Trichodesmium* Tery_3377 gene was integrated at the *psbA2* locus in both the ΔfutA1 and ΔfutA2 mutants such that the gene’s expression is under the control of the *psbA2* promoter. PCR confirmed that the strains (ΔfutA1+ and ΔfutA2+) were homozygous for the *Trichodesmium* gene (data not shown) and sequencing (Eurofins MWG Operon, Ebersberg, Germany) confirmed no errors.

Expression of FutA1, FutA2 and Tery_3377 in the WT, mutants and complemented strains (as applicable) under normal (Fe+) and Fe-limited (Fe-) growth was determined through immunoblotting with anti-IdiA antibodies (for detection of FutA1, FutA2 and Tery_3377) or anti-Slr0513 (only cross reacts with FutA2) (Michel & Pistorius 1992; Tölle et al. 2002) (Fig. 4.9). Due to instability and poor growth of the ΔfutA1/futA2 double mutant, this and the its corresponding, Tery_3377 expressing strain, ΔfutA1/futA2+, were omitted from subsequent analysis.

**Figure 4.9: Expression of FutA1 and FutA2 in *Synechocystis* deletion and complemented mutants.** Antibodies anti-IdiA (top) and anti-Slr0513 (bottom), raised against the *Synechococcus* and *Synechocystis* corresponding homologues respectively (Michel & Pistorius 1992; Tölle et al. 2002), identified the correct combination of FutA1 (blue arrow), FutA2 (green arrow) and Tery_3377 (red arrow) proteins in the WT, all constructed futA deletion and Tery_3377 expressing strains (1-7). Protein extracts were acquired during Fe-deplete growth for the WT and deletion mutants (1-4) and Fe-replete growth for the complemented strains (5-7).
To identify whether the periplasmic localisation of Tery_3377 (Sec.4.2.2 and 4.2.3) is associated with a function in Fe uptake, a phenotypic experiment was designed utilising the high affinity $\text{Fe}^{2+}$ ligand ferrozine (FZ). The ligand which makes $\text{Fe}^{2+}$ inaccessible as it cannot penetrate the cell membrane (Kustka et al. 2005; Shaked et al. 2005; Garg et al. 2007) was previously used in *Synechocystis* growth experiments to suggest that Fe reduction precedes transport across the cell membrane (Kranzler et al. 2011). Here, the ligand was hypothesised to cause an Fe-stress phenotype through reduction of *Synechocystis* Fe uptake, particularly in the case of $\Delta$mutA2 which is presumably already compromised in Fe$^{2+}$ uptake.

During exponential growth, $\Delta$mutA2 experienced a reduction in $F_v/F_m$ when FZ (200 $\mu$M) was included in BG11 media (FZ+) compared to when omitted (FZ-). This was a larger decrease (1.7 times) than the one observed between FZ-/FZ+ conditions in the WT (Fig. 4.10A). Conversely, upon addition of FZ, $\Delta$mutA1 did not experience a significant decrease in $F_v/F_m$. Additionally, compared to the WT, $\Delta$mutA1 had higher $F_v/F_m$ both under FZ+ and FZ- conditions (Fig. 4.10A). The apparent improved $F_v/F_m$ state of $\Delta$mutA1 could be explained by previous observations of increased Fe reduction and uptake by $\Delta$mutA1 compared to $\Delta$mutA2 and WT (Kranzler et al. 2014).

Contrariwise, the $\Delta$mutA1 functional absorption cross section of PSII ($\sigma^{\text{PSII}}$) experienced a marked increase (16%) in FZ+ compared to FZ- (Fig. 4.10B). This was larger than the observed increase in $\Delta$mutA2 (11%) under the same conditions, while additions of FZ to the WT did not induce a statistically significant change (Fig. 4.10B). It is possible that the increase in $\sigma^{\text{PSII}}$ under conditions of Fe-stress is stimulated by an accumulation of phycobilisome (Pcb) antenna complexes surrounding PSII (Bibby et al. 2003; Chen & Bibby 2005). An increase in the light harvesting efficiency of PSII could in that scenario counterbalance a potential reduction of the Fe containing photosystem. As the $\Delta$mutA1 is hypothesised to lack PSII protection from Fe-stress generated ROS, the PSII subunits might be subjected to increased damage in this mutant. Therefore, it can be hypothesised that an increase of $\sigma^{\text{PSII}}$ is caused by upregulation of Pcb stimulated to enhance the efficiency of the particularly compromised PSII subunits in $\Delta$mutA1 under Fe-deficient conditions in FZ+. The larger overall $\sigma^{\text{PSII}}$ of $\Delta$mutA2 (Fig. 4.10B) (albeit its milder response to FZ-additions), could be indicative of a more pronounced state of cellular Fe and oxidative stress.
Figure 4.10: Photosynthetic physiology of strains under Fe stress. *Synechocystis futA* mutants before (ΔfutA1 and ΔfutA2) and after (ΔfutA1+ and ΔfutA2+) insertions of the *Trichodesmium Tery_3377* were compared to the WT when grown in YBG11 media with (green) and without (blue) the Fe^{2+} binding ligand ferrozine (FZ). The photosynthetic physiology during exponential growth was recorded as $F_v/F_m$ (A) and $\sigma^{PSII}$ (B). Statistical significance of differences observed was assessed using a GLM followed by a post-hoc Tukey-test (CI=95%). Between-strain (dissimilar letter group: x,y and z for Fz- and a,b and c for FZ+), and within-strain (*), differences (P < 0.05) are indicated. Error bars show the standard deviation from the mean of three biological replicates.
In addition to photosynthetic physiology, growth rate changes also occurred. In FZ- the growth rate was lower for $\Delta\text{futA1}$ (0.39 day$^{-1}$) compared to the WT (0.45 day$^{-1}$) and $\Delta\text{futA2}$ (0.43 day$^{-1}$). Reductions were observed in all three strains when FZ was added to the media (Fig. 4.11). This reduction was higher (2.5 times) for $\Delta\text{futA2}$ than the $\Delta\text{futA1}$ and WT.

**Figure 4.11: Growth rate of strains under Fe stress.** Mutants $\Delta\text{futA1}$ and $\Delta\text{futA2}$ and their Tery_3377 expressing equivalents $\Delta\text{futA1}+$ and $\Delta\text{futA2}+$ were compared to the *Synechocystis* WT in YBG11 media with (green) and without (blue) ferrozine (FZ). Statistical significance, tested with a GLM followed by a post-hoc Tukey-test (CI=95%), is indicated between (dissimilar letter group: x,y and z for FZ- and a,b and c for FZ+) and within (*) strains.

To summarise, inclusion of FZ in the media seemed to exacerbate the Fe-stress phenotype of strains and this difference was more prominent for $\Delta\text{futA2}$ compared to the WT and $\Delta\text{futA1}$ supporting our hypothesis that stress induced by Fe$^{2+}$ uptake inhibition is more acute in the Fe$^{3+}$ uptake compromised $\Delta\text{futA2}$. Data presented also suggests that the inferred increase in Fe uptake (Kranzler et al. 2014) could potentially enhance $\Delta\text{futA1}$ photosynthetic efficiency above the baseline WT levels.

Involvement of the *Trichodesmium* Tery_3377 in Fe$^{3+}$ uptake and ability to replace FutA2 as part of a functional ABC transporter is hypothesised to cause restored growth and photosynthetic physiology in the $\Delta\text{futA2}+$ strain compared to the $\Delta\text{futA2}$ mutant. Indeed, contrasting the phenotype of $\Delta\text{futA2}$, the $F_v/F_m$, $\sigma^{\text{PSII}}$ and growth rate of strain $\Delta\text{futA2}+$, in FZ+, are no different to those of the WT. (Fig. 4.10, 4.11). In addition, the physiological changes ($F_v/F_m$, $\sigma^{\text{PSII}}$ or growth rate) observed between FZ- and FZ+ growth conditions in
ΔfutA2 are not recoded in ΔfutA2+ indicating independence of its function to Fe$^{2+}$ availability at the growth conditions used here (Fig. 4.10, 4.11).

The phenotype of ΔfutA1+ mimicked that of ΔfutA2+ in all parameters ($F_v/F_m$, σ$^{PSII}$ and growth rate) (Fig. 4.10, 4.11), possibly indicating that a high influx of Fe$^{3+}$ through the activity of Tery_3377 sets the function of FutA1 (protection of PSII under Fe-stress) redundant. However, it cannot be excluded that a secondary intracellular function of Tery_3377 is complementing the ΔfutA1 mutant phenotypes as well.

The small decrease in $F_v/F_m$ and growth rate of ΔfutA1+ and ΔfutA2+ compared to the WT might be the effect of an energetic cost of Tery_3377 overexpression from the psbA2 promoter.

4.2.6 Protein purification and crystallisation

To further confirm the Fe$^{3+}$ binding/transporting properties of Tery_3377, the protein was hexa-histidine-tagged, over expressed in E.coli, extracted and purified to be analysed using crystallographic studies.

After induction with Isopropyl β-D-1-thiogalactopyranoside (IPTG), E. coli carrying the Tery_3377 containing pET21a_3377 plasmid was maintained for 3 hours at 37 °C or overnight at 18 °C. Both batches were subsequently processed and the protein was purified by immobilised metal affinity chromatography (IMAC) on a nickel column. SDS-PAGE and Coomassie staining indicated higher protein yields when E.coli was grown at 18 °C overnight compared to 3 hours at 37 °C (data not shown). Protein eluted from the IMAC column was pooled and further purified by size exclusion chromatography (SEC), the elution profile of which (Fig. 4.12) was characterised by one prominent and one secondary peak. Through SDS-PAGE of eluted samples, Tery_3377 was identified as the protein responsible for the prominent peak.

The red colouration of the purified protein (spin concentrated to 42 mg/ml) indicated the protein’s Fe binding ability (Fig. 4.13). Crystal trials were set using protein before or after Fe (FeCl$_3$)-loading to acquire the apo and fully loaded structures respectively. Crystals formed after a week (Fig. 4.13) and the structure is currently being resolved (in collaboration with Moritz Machelett and Ivo Tews, Institute for Life Sciences, University of Southampton). Preliminary data indicate the Fe$^{3+}$ binding capacity of Tery_3377 at the structural level (data not shown).
Figure 4.12: Size Exclusion Chromatography (SEC) of Tery_3377. (A) The purification was performed on a HiLoad 16/60 Superdex 200 size exclusion column. Fractions from peaks 1 and 2 as indicated were visualised through SDS-PAGE and Coomassie staining (B). Peak 2 protein samples were the correct size (36 kDa) for recombinant Tery_3377, thus fractions 31-34 were used in subsequent analysis.
Figure 4.13: Purification and crystallisation of Tery_3377. The protein was overexpressed in E.coli which was induced with IPTG and kept for 16 hours at 18 °C. The cells were harvested and the hexa-histidine-tagged protein was extracted and subsequently purified through Ni\(^{2+}\)-affinity chromatography and size-exclusion chromatography (SEC) to give a coral-red coloured protein eluate (A). The protein crystallised into the apo-form (B) and Fe-loaded form (C) and the structure is currently being resolved.
4.3 Conclusions and further discussion

In an ocean of nutrient stress, primary producers have developed a plethora of adaptive strategies to survive and flourish. Two of these strategies are exemplified by the reported functions of a single protein which had been demonstrated to either contribute to Fe$^{3+}$ uptake or protection of PSII under oxidative stress. Homologues of this protein have been differently annotated (FbpA, FutA, IdiA, AfuA, SfuA) according to their function or the organism they have been characterised in. To avoid confusion, I propose the conservation of the existing names, with FbpA assigned to the homologue of all pathogenic bacteria and the separation of cyanobacterial paralogues in FutA2 for the transporter binding protein and IdiA (FutA1) for the intracellular protein.

Bioinformatic analysis of the *Trichodesmium* homologue (Tery_3377), currently annotated as IdiA, indicates a closer similarity to FutA1 contradicting the common assumption that this protein is an Fe$^{3+}$ transporter. However, GFP fusion to the protein signal sequence and heterologous expression in the model organism *Synechocystis* PCC 6308, indicates a clear periplasmic localisation (Sec. 4.2.2). Site directed mutagenesis (SDM) reveals that the mechanism for its export utilises the Twin Arginine Translocase (TAT) system (Sec. 4.2.3). In the periplasm Tery_3377 binds Fe (Sec. 4.2.6) and appears to contribute to cellular Fe$^{2+}$ uptake (Sec. 4.2.5). Based on the results presented in this Chapter, Tery_3377 should be re-annotated as FutA2 instead of IdiA, as through the latter an intracellular function similar to the corresponding *Synechococcus* IdiA protein is inferred. Following these results, the cellular need for an alternative PSII protection strategy remains to be examined. The possibility of a Tery_3377 secondary intracellular role cannot be excluded (based on its complementation of the *Synechocystis futA1* mutant) (Sec. 4.2.5), and would be interesting to further address this question in the future.

A better characterisation of the two cyanobacterial variants is necessary to understanding potential secondary functions of Tery_3377. Contradicting results have previously been published regarding the two paralogues’ localisation and function. Currently understood to bind Fe$^{3+}$ (Katoh et al. 2001b; Waldron et al. 2007; Badarau et al. 2008), the two variants have previously been reported to carry Fe$^{2+}$ (Koropatkin et al. 2007) and both have been assumed to be involved in Fe transport in early studies (Katoh et al. 2001b). FutA2 is now established as an Fe$^{3+}$ transporter, due to the detrimental effect of its deletion to Fe uptake, and its periplasmic localisation (Fulda et al. 2000; Durán et al. 2004; Waldron et al. 2007; Badarau et al. 2008). However, it also potentially contributes to other processes such
as the reduction of Fe$^{3+}$ during Fe uptake (Kranzler et al. 2014), or binding of periplasmic Fe to prevent obstruction of other metal transporters (e.g. copper) (Waldron et al. 2007).

The function of IdiA (FutA1) is even less well characterised than that of FutA2. In *Synechocystis*, ΔfutA1 mutants have been demonstrated either exhibit increased (Kranzler et al. 2014) or decreased (Katoh et al. 2001b) Fe uptake efficiency compared to the WT. For the former, strains of *Synechocystis* used are not derived from an isogenic parent, something that raises the risk of diverging irrelevant mutations that influence the study results. However, hypothesising that this is not the case, the observations could be explained by an upregulation of futA2 and feoB expression and consequently Fe uptake (Kranzler et al. 2014) as a way of responding to the loss of FutA1 function in ΔfutA1.

Here the presence of a valid FutA2 signal sequence acting to direct the protein to the periplasm is confirmed. Interestingly, traces of FutA1 signal sequence periplasmic expression are also observed even though we cannot exclude the case of unspecific translocation of the protein due to its potentially high levels of expression (expressed under the control of the *psbA2* promoter). From microscopic images alone this does not appear to be the case (no high intracellular pools of GFP are evident) but due to the lack of quantitative data on GFP expression we are unable to support this argument. Further study is required to identify whether FutA1 (IdiA) has an additional periplasmic role, possibly involved in Fe uptake.

To conclude, in this Chapter, the identity of Tery_3377 as an Fe$^{3+}$ transporter is supported and its reannotation to FutA2 is suggested. However, further work is required to fully characterise the cyanobacterial FutA2 and IdiA paralogues and identify possible secondary roles of Tery_3377 in *Trichodesmium*’s Fe-stress metabolic response.
Chapter 5
Desert dust as a source of iron-
A physiological and transcriptomic study of Trichodesmium

5.1 Introduction

The main source of iron (Fe) to the open ocean is Aeolian dust deposition from the world deserts (Jickells et al. 2005; Ussher et al. 2013). It is expected that at circumneutral pH and oxic-conditions encountered in the surface oceans, Fe in dust (present as oxides, hydroxides and in aluminosillicate minerals) will in its majority be maintained as particulate (> 0.4 μm) or colloidal (0.02-0.4 μm) instead of soluble (< 0.02 μm). This is supported by laboratory experiments which suggest that within the dissolved fraction (< 0.2 μm) colloidal as opposed to soluble is the dominant form of dust Fe (Aguilar-Islas et al. 2010; Fitzsimmons et al. 2015). In addition, comparisons between computations of the total dust inputs and the Fe inventory of the oceans indicate solubility of dust Fe to be minimal (~1-2%) (Jickells & Spokes 2001).

The primary form of Fe utilised by phytoplankton is typically assumed to be soluble Fe and until recently it was believed that particulate/colloidal Fe is largely unavailable to these organisms. However, a growing number of studies is demonstrating the opposite (Nodwell & Price 2001; Frew et al. 2006; van der Merwe et al. 2015). Bioavailability seems to be not solely dependent on the speciation and solubility of Fe, with the identity of phytoplankton and the molecular tools in their repertoire proving critical.
Chapter 5: Dust as a source of iron to \textit{Trichodesmium}

Amongst their Fe acquisition strategies, phytoplankton employ siderophores, chelating agents with variable structures that are biotically produced and released in their immediate environment. There, they compete with other natural ligands and bind Fe\textsuperscript{3+} increasing its bioavailability (Neilands 1995). Moreover, microbially produced ligands also facilitate photolysis of Fe\textsuperscript{3+} to Fe\textsuperscript{2+} in the ocean’s surface layer (Barbeau et al. 2001). The majority of known siderophores are produced by heterotrophic bacteria (Hopkinson & Morel 2009) with \textit{Synechococcus} (Ito & Butler 2005; Boiteau & Repeta 2015) being the only marine cyanobacteria that have been identified to produce such ligands. However, other microorganisms (Poole et al. 1990; Guan et al. 2001) and amongst them \textit{Trichodesmium} (Achilles et al. 2003; Roe et al. 2012), possess the ability to engage in a form of ‘siderophore piracy’ and benefit through the uptake of exogenous siderophores.

Uptake of organically complexed Fe can be direct or involve an extracellular reduction step and consequent release from its ligands before transport. In addition to organically complexed Fe species, reductive mechanisms possibly also target particulate and colloidal inorganic Fe (Rubin et al. 2011). Studies using the ligand ferrozine (which strongly binds Fe\textsuperscript{2+} and renders it unavailable for uptake) indicate a requirement for Fe reduction prior to its uptake by various phytoplankton species including eukaryotes as well as cyanobacteria such as \textit{Synechocystis} and \textit{Trichodesmium} (Shaked et al. 2005; Lis & Shaked 2009; Kranzler et al. 2011; Roe & Barbeau 2014).

Eukaryotic phytoplankton possess transmembrane ferric reductases (FRE) which transfer electrons to complexed Fe\textsuperscript{3+} and reduce it to Fe\textsuperscript{2+} facilitating its dissolution (Kustka et al. 2007). Consequently, the elevated Fe\textsuperscript{2+} concentrations surrounding the cell create potential for its uptake through divalent transporters including ZIP and NRAMPs (Kustka et al. 2007; Allen et al. 2008; Blaby-Haas & Merchant 2012). Further, multi-copper oxidases (MCO) (with similarity to the yeast multicopper oxidase, FET), can reoxidise Fe\textsuperscript{2+} to Fe\textsuperscript{3+} before uptake through associated Fe\textsuperscript{3+} permeases (FTR) (Herbik et al. 2002; Boukhalfa & Crumbliss 2002; Maldonado et al. 2006; Y. Paz et al. 2007a). Recently, a copper oxidase-independent Fe\textsuperscript{3+} uptake system was also identified to function in eukaryotic phytoplankton through the membrane protein ISIP2a (Morrissey et al. 2015).

Reduction and uptake of Fe by cyanobacterial cells is less well understood and the components involved are yet to be fully elucidated. In \textit{Synechocystis} PCC 6803 \textit{Synechocystis} hereafter), the transfer of electrons to Fe\textsuperscript{3+} was suggested to be mediated through the plasma-membrane-located alternative respiratory terminal oxidase (ARTO),
Chapter 5: Dust as a source of iron to *Trichodesmium*

with the knockout mutant displaying impaired Fe reduction and uptake (Kranzler et al. 2014). Also, deletions of pili (Lamb et al. 2014), structures known to be electrically conductive (Reguera et al. 2005; Gorby et al. 2006) inhibit the organism’s growth on Fe oxides.

Moreover, the export of reactive oxygen species (ROS) has been proposed to contribute to reductive Fe uptake in both eukaryotic and prokaryotic phytoplankton. Several species including *Trichodesmium* were previously shown to produce ROS (Henry & Vignais 1980; Yamasaki et al. 2004; Marshall et al. 2005; Rose et al. 2005; Godrant et al. 2009; Roe & Barbeau 2014; Hansel et al. 2016) and evidence suggests that superoxide can contribute to inorganic Fe (but not to ferric citrate) uptake in *Trichodesmium* (Roe & Barbeau 2014).

Once reduced, Fe can very quickly revert to its original oxidised state, therefore binding of Fe$^{2+}$ and microbial adaptations that increase its stability, as well as its efficient transport, are necessary. The puff colony morphotype of *Trichodesmium* was hypothesised to increase uptake of Fe from dust by maintaining an anoxic colony core environment, favourable in keeping Fe$^{2+}$ reduced (Hall & Pearl 2011; Rubin et al. 2011). Environmentally collected *Trichodesmium* puff colonies attach to desert dust particles and coordinate in an incredible example of cooperative multicellularity to centralise dust to the colony cores (Rubin et al. 2011). Previous experiments have also recorded dust-stimulated colony formation by *Trichodesmium erythraeum* IMS101 in culture (Langlois et al. 2012).

Although processes associated with the *Trichodesmium* cell surface have been suggested to enhance dissolution of Fe from desert dust and ferrihydrite, Rubin et al. 2011 only observe this to be the case for *Trichodesmium* puff colonies and not filamentous *Trichodesmium* in culture. The latter could enhance dissolution from ferrihydrite, approximately 13 x less efficiently than puffs. Nevertheless, filamentous laboratory grown *Trichodesmium* is able to access desert dust-bound Fe for growth (Langlois et al. 2012).

The contribution of the cell surface to Fe dissolution was previously supported (Rubin et al. 2011) through comparisons of Fe uptake from free $^{55}$Fe ferrihydrite to that trapped within the fibres of a glass fibre filter. The possibility that this entrapment blocks dissociation of free inorganic Fe (Fe’) from ferrihydrite particles was not addressed. Also EDTA was not included in the experimental media and therefore re-precipitation of Fe’ following its dissolution is possible when binding by the cells is prevented (Rubin et al. 2011). Therefore, although evidence seems to support that the cellular processes promote dissolution of Fe
Chapter 5: Dust as a source of iron to *Trichodesmium*

From colloidal ferrihydrate, the importance of cell-to-substrate contact for this process is yet to be fully determined.

Using *Trichodesmium erythraeum* IMS101 cultures (hereafter *Trichodesmium*) we attempt to better characterise the Fe uptake strategies involved in accessing Saharan desert dust Fe. Through the development of a simple experimental design, we evaluate the significance of cell surface contact for this process. Furthermore, the organism’s response to dust and Fe-stress are compared at the transcriptomic level by RNA sequencing analysis with the aim of describing the molecular pathways coordinating them.

### 5.2 Experimental design

The importance of cell-to-substrate contact for acquisition of Fe from Saharan desert dust by *Trichodesmium*, was assessed by separating the cells from dust added to the media (mYBC-II), using dialysis tubing ([Dust]). This was compared to cultures in which dust was included in the direct vicinity of the cells (Dust+) and no-dust treatments with (Fe+) or without (Fe-) the addition of 400 nM FeCl$_3$-EDTA. All cultures contained dialysis tubing to keep experimental conditions identical across treatments and a no-dialysis tubing control (no DT) was used to assess potential Fe contamination as a result its inclusion (Fig. 5.1). The experiment was run in triplicate and was maintained for 17 days during which the physiology of *Trichodesmium* was monitored.

![Figure 5.1: The experimental design](image)

**Figure 5.1: The experimental design.** Dust was provided to cultures with mYBC-II and equal concentrations of *Trichodesmium* cells either separated by a dialysis membrane such that leachate but not dust comes in contact with cells ([Dust]) or in the direct vicinity of cells (Dust+) permitting their interaction. Set in the same way cultures with (Fe+) or without (Fe) added FeCl$_3$-EDTA were used for comparison. Dialysis tubing, was included in all treatments to correct for any possible effects, and no-dialysis tubing cultures (no DT) were maintained as a control.
To test whether nutrients from dust dissolved preferentially when dust was free in the culture vessels, compared to when enclosed in dialysis tubing, *Trichodesmium*’s physiological response was monitored in a separate control experiment. This involved the incubation of experimental treatments, identical to the ones set in the abovementioned experiment, but without *Trichodesmium* cells, for a period of 14 days before filtration in clean culture flasks and subsequent inoculation with equal concentrations of cells (Fig. 5.2). The growth of cultures was monitored for a period of 9 days thereafter.

![Experimental design to assess possible differential dissolution of nutrients from dust enclosed or not enclosed in dialysis tubing.](image)

*Figure 5.2: Experimental design to assess possible differential dissolution of nutrients from dust enclosed or not enclosed in dialysis tubing.* Conditions used were identical to the previous experiment but *Trichodesmium* cells were omitted. Following a period of 14 days all conditions were filtered through 0.22 μm sterile filters (Millipore, Watford, UK) and mYBC-II was inoculated with equal concentrations of *Trichodesmium* cells.
5.3 Results and discussion

5.3.1 Experimental controls

Using a control treatment, the effect of washed dialysis tubing (DT) to Fe deficient media was examined. Comparison between the Fe deficient treatment with (Fe-) and without (no DT) DT revealed no statistical differences in $F_v/F_m$ (estimate of the apparent photosystem II photochemical quantum efficiency) and $\sigma_{\text{PSII}}$ (functional absorption cross section of photosystem II) or growth rates and filament lengths indicating no/insignificant trace metal contamination during installation of the DT (Fig. 5.3).

\[ \text{Figure 5.3: Assessment of potential dialysis tubing effects when included in the experiment.} \]

Comparison of the no dialysis tubing (no DT) control treatment to the equivalent Fe- treatment with DT included (Fe-) indicates that growth rate during exponential growth, as well as filament lengths and photosynthetic physiology parameters ($F_v/F_m$ and $\sigma_{\text{PSII}}$) measured on the final day of the experiment, did not differ significantly between the two treatments. Error bars represent standard deviations from the mean of three biological replicates.
The second control experiment was aimed at investigating whether the experimental set up alone promoted differential release of dust-associated chemicals affecting *Trichodesmium* between [Dust] and Dust+ treatments. No statistically different growth was identified in media preincubated with dust inside or outside the DT and growth rates in both cases did not exceed those of the Fe- cultures. On the contrary, in Fe+ media growth rates were significantly (GLM and Tukey test, P < 0.05) faster (25-35%) compared to all other treatments (Fig. 5.4).

**Figure 5.4: Assessment of potential inhibitory effect of the dialysis tubing on abiotic dust-chemical dissolution.** Similar growth rate of *Trichodesmium* in preincubated (p) Dust+ and [Dust] treatments indicates that differential dissolution of nutrients from dust is not affecting the experimental results. The growth rate is represented as the percentage of the highest (recorded in the Fe+ treatment). Distinct letter groups (A and B) mark significant differences (GLM and Tukey test, P < 0.05) between Fe+ and the remaining treatments. Error bars are standard deviations of the mean of triplicate cultures.

### 5.3.2 Dust as a source of iron - the role of cell surface contact

As the experimental design was tested through control experiments and the results excluded major interference and bias by the DT, the set up was justified as a valid approach to address our scientific questions. Therefore, the effect of Saharan desert dust to *Trichodesmium* physiology when in contact (Dust+) or separated ([Dust]) from the cells was compared to Fe- and Fe+ treatments to investigate the potential contribution of cell surface processes to Fe acquisition.
Chapter 5: Dust as a source of iron to *Trichodesmium*

During the course of the experiment growth of *Trichodesmium* was significantly (GLM and Tukey test, P < 0.05) faster in Fe+ (0.35 day\(^{-1}\)) and Dust+ (0.31 day\(^{-1}\)) compared to Fe- (0.17 day\(^{-1}\)) and [Dust] (0.22 day\(^{-1}\)) for which growth rate was 1.8-2 and 1.4-1.6 times reduced respectively (Fig. 5.5). This indicates the preferential growth in the presence of FeCl\(_3\), but also when dust is added to the direct environment of *Trichodesmium* cells. The same does not apply when dust is separated from the cells (significantly slower growth in [Dust] compared to both Dust+ and Fe+).

**Figure 5.5: Growth of *Trichodesmium* under different Fe regimes.** Cell concentration in dust added (Dust+, dark brown), dust separated ([Dust], light brown), FeCl\(_3\) - EDTA added (Fe+, yellow) or Fe deficient (Fe-, green) cultures was monitored over 17 days and **(A)** growth rates were calculated during late exponential phase. Growth rates presented are scaled relative to the maximum (Fe+) and statistically significant differences (GLM and Tukey test; P < 0.05) between treatments are indicated with dissimilar letter groups (A, B). Error bars represent standard deviations of three biological replicates. **(B)** Visual assessment of the cultures reveals different pigment concentration across treatments on the last day of the experiment (day 17).
Furthermore, on the last day of the experiment filaments were on average shorter in Fe+ (0.45 mm) compared to Fe- (0.80 mm) and Dust+ (0.76 mm) cultures (GLM and Tukey test, \( P < 0.05 \)). The filament length of [Dust] cultures (0.72 mm) was not significantly dissimilar to any of the other treatments (Fig. 5.6). Deviation in *Trichodesmium* filament lengths between Fe+ and Fe- conditions was previously described in literature, but this has not always been consistent with some studies reporting longer filaments in Fe+ compared to Fe- (Küpper et al. 2008) and some in Fe- compared to Fe+ (Roe et al. 2012). Iron limitation experiments conducted under our laboratory conditions have consistently supported the results presented here (Snow et al. 2015).

Although the shorter size of Fe+ filaments could be the result of faster growth rate in this treatment, this explanation cannot justify the similarity between Fe- and Dust+. We suggest that a different process might be stimulating persistence of longer filaments in the Dust+ and [Dust] treatments. It can be hypothesised that stimuli from dust-leached minerals trigger this physiology and the possibility that longer filaments are favourable to more efficiently access particulate or colloidal Fe requires further study. Multicellularity has been previously linked to dust processing (Rubin et al. 2011) and studies indicated induction of colony formation by dust in *Trichodesmium* filamentous cultures (Langlois et al. 2012).

The growth rate dissimilarity between treatments is paralleled by difference in photosynthetic efficiency (\( F_v/F_m \)). In Fe- and [Dust] treatments \( F_v/F_m \) declined from day 6 to day 17 by 0.15 and 0.16 respectively while Fe+ cultures established a significantly (GLM and Tukey test, \( P < 0.05 \)) higher \( F_v/F_m \) compared to the rest of the treatments from day 6 onwards (\( F_v/F_m \) on day 6= 0.53). Dust+ cultures experienced an initial drop on day 6 (0.42) with recovery by day 8 (0.5) and remained significantly (GLM and Tukey test, \( P < 0.05 \)) higher than Fe- and [Dust] cultures until the end of the experiment (Fig. 5.7). The differences in \( F_v/F_m \), if used as an indication of cellular photosynthetic activity, provide justification to the distinct growth rates observed between treatments.
Figure 5.6: *Trichodesmium* filament physiology under different Fe regimes. FeCl₃-EDTA added (Fe+) cultures (A: yellow, B: top right) had significantly (GLM and Tukey test; P < 0.05) shorter filaments compared to Fe deficient (Fe-) (A: green, B: top left) and dust added (Dust+) cultures (A: dark brown, B: bottom right). In the dust separated ([Dust]) treatment (A: light brown, B: bottom left) filaments were not different in size to any other treatment. Statistically different filament lengths are indicated by distinct letter groups (A, B). Error bars are standard deviations of triplicates.
Chapter 5: Dust as a source of iron to *Trichodesmium*

Figure 5.7: *Trichodesmium* photosynthetic efficiency ($F_v/F_m$) under different Fe regimes. Samples from Fe+ (yellow), Dust+ (dark brown), Fe-(green) and [Dust] (light brown) cultures were taken after 2-2:30 hrs of light during a 12 hr: 12 hr light: dark cycle for 17 days of growth. Statistically different groupings (GLM and Tukey test; $P < 0.05$) between treatments for each day are marked (dashed squares) until the end of the experiment (sampling day, SD). Error bars are standard deviations of the means of three biological replicates for each treatment.

The functional absorption cross section of PSII (PSII), $\sigma^{\text{PSII}}$, had a more varied response over time and between treatments compared to $F_v/F_m$. The Fe+ treatment had lower values ($0.40$-$0.42$ Å$^2$ quantum$^{-1}$) than the rest ($0.42$-$0.50$ Å$^2$ quantum$^{-1}$) between days 6 to 14 and this difference was on some occasions significant (GLM and Tukey test, $P < 0.05$) (Fig. 5.8). By day 17 of the experiment $\sigma^{\text{PSII}}$ of all treatments converges around $0.45$ Å$^2$ quantum$^{-1}$ except for the Dust+ cultures which maintain high values ($0.49$ Å$^2$ quantum$^{-1}$). Iron stress conditions are known to stimulate an increase of $\sigma^{\text{PSII}}$ possibly as a result of accumulating antennae complexes (Küpper et al. 2008). This physiological behaviour, currently not fully understood, is thought to perhaps allow a reduction in the number of the Fe expensive photosystem II (PSII) units as does the iron stress induced protein A, IsiA (Bibby et al. 2001) for photosystem I (PSI). Although this could justify the lower $\sigma^{\text{PSII}}$ recorded in the Fe+ compared to the Fe- treatment the same relationship is not identifiable in Dust+ compared to [Dust]. Instead Dust+ $\sigma^{\text{PSII}}$ is maintained at a significantly (GLM and Tukey test, $P < 0.05$) higher level compare to that of Fe+ cultures (day 6 -17). Shading or reflection of light by dust particles has to be considered as a possible driver for bigger antenna size requirements in Dust+. 
Figure 5.8: *Trichodesmium* functional absorption cross section of PSII ($\sigma_{\text{PSII}}$) during growth under different Fe regimes. (A) The treatments Fe+ (yellow), Dust+ (dark brown), Fe- (green) and [Dust] (light brown) are compared to each other on different days of the experiment and (B) changes in their $\sigma_{\text{PSII}}$ are assessed over time. Statistical significant differences in A (GLM and Tukey test; $P < 0.05$) is indicated by a distinct letter group (a, b). Error bars are the standard deviations from the mean of three replicate cultures.
Looking at the evolution of the $\sigma_{\text{PSII}}$ response in each treatment over time allows the identification of two distinct patterns of behaviour amongst the four treatments (Fig. 5.8B). The first is observed in the [Dust] and Fe- treatments as an elevated $\sigma_{\text{PSII}}$ on days 10 and 14 compared to days 6 and 17 creating an inverted U-shape pattern. The late reduction of the photosystem antennae on day 17 could be due to an extreme exhaustion of Fe resources beyond the threshold at which an increase would be beneficial to the cells. The second behaviour is observed in the Fe+ and Dust+ treatments and involves the steady increase or maintenance of $\sigma_{\text{PSII}}$ over time (day 6-17). It is possible that this describes the fine tuning of the antennae as light or Fe concentrations in the media are decreasing in the Fe+ and Dust+ treatments over time.

The deviation of physiological parameters, including growth rate and $F_v/F_m$ as well as the $\sigma_{\text{PSII}}$, between Dust+ and [Dust] treatments, indicates an advantage when dust particles are in direct physical contact with cells. These differences resemble those measured between Fe- and Fe+ suggesting differential Fe uptake efficiency in the two dust treatments. The observations provide the first direct evidence for a link between cell growth and photosynthetic physiology to cell surface contact with dust.

5.3.3 RNA sequencing results- An overview

To determine the underlying mechanisms resulting in the physiological differences between Dust+ and [Dust] and compare them to those between Fe+ and Fe-, samples from all conditions were acquired on the last day of the experiment (day 17) at 5 hrs through the light period (12 hr: 12 hr light: dark cycle) and RNA was extracted and sequenced. Reads were aligned against the *Trichodesmium erythraeum* IMS101 genome which is composed of 4451 annotated genes only ~54% of which are annotated with a Gene Ontology (GO) classification number (GO annotated) (Fig. 5.9). Expression of the majority of *Trichodesmium* genes (4342 genes) was identified in at least one treatment and 4076 were identified in all 4 treatments. Of these 3257 genes had read counts above the statistical threshold required for inclusion in differential gene expression analysis, as determined by the DESeq2 Bioconductor package (Love et al. 2014). Expression of approximately one third of those genes (1050) was significantly affected by growth condition ($P < 0.05$, DESeq ANODEV) (Fig. 5.9). It can also be observed that greater statistical power for detection leads to a higher identification of differentially expressed genes at increased expression intensities (aligned read counts) (Fig. 5.10).
Chapter 5: Dust as a source of iron to *Trichodesmium*

Figure 5.9: Gene numbers and GO annotation coverage. The majority of *Trichodesmium* genes (Genes) are identified in at least one condition of the experiment (Identified genes). Genes for which expression is recorded in all treatments above threshold quantities (Genes included in analysis) are analysed to identify differential expression (differentially expressed genes). A large proportion of the *Trichodesmium* genome is not GO annotated (NA, light grey).

![Gene numbers and GO annotation coverage](image)

Figure 5.10: Distribution of the number of genes expressed across different levels of reads/gene. The distribution of the number of genes identified in the analysis (light grey) is compared to the distribution of the number of differentially expressed genes (dark grey). Detection of differential expression is skewed towards highly expressed genes.

![Distribution of the number of genes expressed across different levels of reads/gene](image)
The validity of replicates within each treatment was confirmed through clustering analysis which showed closer intra-treatment to inter-treatment similarities. Treatments clustered in two distinct sets with closer similarity between [Dust] and Fe- cultures and between Fe+ and Dust+ (Fig. 5.11). This indicates that the growth conditions stimulate related gene expression phenotypes between those groups in support of growth and photosynthetic physiological responses observed (Sec. 5.3.2).

In addition, comparisons of the numbers of genes differentially expressed in treatments Dust+, [Dust] and Fe- relative to Fe+ is presented in Figure 5.12. The highest dissimilarity to Fe+ was identified in [Dust] samples (797 genes), while 419 genes are differentially expressed compared to Dust+, and 438 genes compared to Fe-. Increased changes in gene expression between of Fe+ and [Dust] is perhaps the result of no access to Fe in this treatment (compared to Dust+) but at the same time dissolution of dust-bound chemicals, altering the chemical composition of the media (as opposed to Fe-).

**Figure 5.11: Gene expression clustering across treatments and replicates.** Dendrogram from average linkage clustering of treatments [Dust], Fe-, Dust+ and Fe+ indicates that replicates of each treatment are grouped together while two main clusters (1, 2) describe closer expression relationships between treatments [Dust] and Fe- (1). and Dust+ with Fe+ (2).

When compared to gene expression in Fe+ conditions, Dust+ and [Dust] treatments share 125 genes expressed similarly, [Dust] and Fe- 179 genes, while Fe- and Dust+ only 16 genes. This observation suggests the presence of two main drivers shaping the transcriptomic responses identified in this experiment: (A) the availability of Fe and (B) the presumed
presence of dust-leached chemical stimuli. As identified in the clustering analysis (Fig. 5.11), the former appears to be the overriding driver (Fe+ and Dust+ are more similar than Fe+ and Fe-).

Figure 5.12: Comparison of differential expression across the experimental treatments.
The number of genes differentially expressed uniquely and in combination in the Fe- (green), Dust+ (dark brown) and [Dust] (light brown) treatments is compared to the Fe+ (yellow) treatment.

To visualise the array of expression patterns across the experimental treatments (Fig. A.1) and characterise the drivers shaping these, the differentially expressed genes were grouped in 10 clusters based on their expression profile across the four experimental conditions (Fig. 5.13, Table 5A). The Dust+ stimulated genes are observed to group together in Cluster 3 (n=86) while dust leached chemical stimuli seem to affect the expression of genes in Clusters 1 (upregulation in [Dust] and Dust+ compared to Fe+ and Fe-; n=71) and 8 (downregulation in [Dust] and Dust+ compared to Fe+ and Fe-; n=74). The majority of genes fall in Clusters 10 (n=259) and 2 (n=185) in which Fe+ and Dust+ treatments are either experiencing higher (Cluster 2) or lower (Cluster 10) expression levels relative to Fe- and [Dust]. The third (Cluster 6, n=135) and fourth (Cluster 4, n=109) largest clusters represent the bidirectional gene regulation of Fe+ distinctly to Fe-, Dust+ and [Dust] treatments.

These clustering results indicate that the majority of genes follow a linked response in Fe+ and Dust+ compared to [Dust] and Fe- but also a considerable number of genes in Fe+ but not Dust+ is uniquely different to the Fe deficient [Dust] and Fe-.

These results are interesting as they point to a response specific to either (1) increased available Fe/higher Fe concentrations, (2) higher presence of Fe as dissolved inorganic Fe’ as opposed to bound in particulate/colloidal dust or (3) dissimilar nutrient composition of the media when Fe is provided alone (Fe+) compared to when provided as dust (Dust+).
Chapter 5: Dust as a source of iron to Trichodesmium

Figure 5.13: Clustering demonstrates patterns of differential expression across all genes and treatments. Ten clusters of different expression profiles across the [Dust] (light brown), Dust+ (dark brown), Fe- (green) and Fe+ (yellow) treatments were generated (Pearson product moment correlation coefficient between each pair of differentially expressed genes and WPGMA). For plots Cluster 1-10 the average expression for each treatment is corrected to the average across all treatments. A dendrogram of the clusters details the expression across all differentially expressed genes in a heatmap.
Table 5A: Clustering of differentially expressed genes in 10 distinct patterns across experimental treatments.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Number of genes</th>
<th>Description</th>
<th>Gene examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>71</td>
<td>Upregulation in [Dust] and Dust+ compared to Fe- and FeCl₃</td>
<td>\text{psbH (Tery_2868)} \text{psbJ (Tery_3507)} \text{50S ribosomal protein L25 (Tery_3283)} \text{cytochrome c (Tery_2561)} \text{hemolysin-type calcium-binding (Tery_0419)} \text{cell surface protein (Tery_0849)}</td>
</tr>
<tr>
<td>2</td>
<td>185</td>
<td>Upregulation in FeCl₃ and Dust+ compared to [Dust] and Fe-</td>
<td>\text{F0F1 ATP synthase subunit C (Tery_2203)} \text{phycobilisome protein (Tery_0983)} \text{phycocyanin, alpha subunit (Tery_5048)}</td>
</tr>
<tr>
<td>3</td>
<td>86</td>
<td>Upregulation in Dust+ compared to FeCl₃ [Dust] and Fe-</td>
<td>\text{ATPase AAA-2 (Tery_2437)} \text{peptidyl-prolyl cis-trans isomerase, cyclophilin type (Tery_2705)} \text{UDP-glucose 6-dehydrogenase (Tery_2477)}</td>
</tr>
<tr>
<td>4</td>
<td>109</td>
<td>Downregulation in FeCl₃ compared to Dust+, [Dust] and Fe-</td>
<td>\text{hemolysin-type calcium-binding (Tery_3467)} \text{tetratricopeptide TPR_2 (Tery_5048)} \text{ATPase (Tery_2501)}</td>
</tr>
<tr>
<td>5</td>
<td>93</td>
<td>Upregulation in Fe- and FeCl₃ compared to Dust+ and [Dust]</td>
<td>\text{ferredoxin-nitrite reductase (Tery_1068)} \text{OpcA protein (Tery_0685)} \text{6-phosphogluconate dehydrogenase (Tery_1834)}</td>
</tr>
<tr>
<td>6</td>
<td>135</td>
<td>Upregulation in FeCl₃ compared to Dust+, [Dust] and Fe-</td>
<td>\text{cytochrome-c oxidase (Tery_0277)} \text{peptidoglycan glycosyltransferase (Tery_2090)} \text{fructose-1,6-bisphosphatase (Tery_0682)}</td>
</tr>
<tr>
<td>7</td>
<td>27</td>
<td>Upregulation in Fe- compared to Dust+, [Dust] and FeCl₃</td>
<td>\text{signal transduction protein (Tery_2045)} \text{ribonuclease III (Tery_2839)} \text{Chk2 histidine kinase (Tery_3912)}</td>
</tr>
<tr>
<td>8</td>
<td>64</td>
<td>Downregulation in Dust+ and [Dust] (more intensely in the former) compared to Fe- and FeCl₃</td>
<td>\text{plastocyanin (Tery_2563)} \text{hemolysin-type calcium-binding region (Tery_2055)} \text{Mo-dependent nitrogenase-like (Tery_4114)}</td>
</tr>
</tbody>
</table>
5.3.4 RNA sequencing technique validation

The results of the RNA sequencing (RNAseq) analysis were validated via quantitative reverse transcription (qPCR) using three of the identified differently expressed genes (Tery_2561, Tery_1667, Tery_4114). The cytochrome c_{553} gene (cyt c_{553}, Tery_2561) was upregulated in Fe+ and Dust+ compared to Fe- and [Dust], iron-stressed-induced protein-A gene (isiA, Tery_1667) downregulated in Fe+ compared to the rest of the treatments and the Mo-dependent nitrogenase-like-protein gene (nif-like, Tery_4114) upregulated in the [Dust] and Fe- compared to Dust+ and Fe+ treatments.

For qPCR normalisation, candidate reference genes rnpB (Tery_R0021) and glyA (Tery_2660), previously used to correct gene expression in Fe-stress experiments (Chappell & Webb 2010), and genes for the gas vesicle protein (gvpN, Tery_2329) and ribonuclease T2 (Tery_3279) selected from stably expressed genes across the experimental treatments (as identified from this RNAseq analysis), were tested. The latter pair was the most stable combination (geNorm, qBase, Biogazelle, Belgium) and was consequently chosen for normalisation of qPCR derived expression values.

Following normalisation, the results were compared to the RNAseq expression of the chosen genes. Patterns of expression from the two methods closely resembled each other and regression analysis indicated a good correlation ($R^2 = 99\%$) (Fig. 5.14).
Figure 5.14: Validation of the RNAseq results using quantitative RT-PCR (qPCR). (A) Regression analysis indicated a correspondence of calculated relative fold changes (RFC) in expression between RNAseq analysis (green) and qPCR (blue) for genes: (B) Tery_2561, (C) Tery_1667, and (D) Tery_4114. RFC is calculated by normalisation to the lowest expression across the four treatments (Fe-, Fe+, [Dust], Dust+).
5.3.5 Differentially regulated processes and genes

As a first step to identifying the nature of changes observed in the transcriptome, gene ontology (GO) classification was utilised. GO classifications stem under the domains of Cellular component, Molecular function and Biological process. However, a large proportion of Trichodesmium genes are not GO annotated with 46% of the differentially expressed genes not belonging to any GO category (Fig. 5.9) and therefore, it is possible that processes and genes less well characterised are missed in the descriptions that follow.

To identify significantly overrepresented GO categories within the differentially regulated genes, singular enrichment analysis (SEA) was performed (AGRIGO, Du et al. 2010) (Fig. 5.15). In the domain of Biological Processes, photosynthesis, including both electron transport and the light reaction of photosynthesis, is overrepresented. Confirming this, genes related to the thylakoid membranes belonging to categories including the light harvesting complex, photosystems I and II, as well as the ATP synthase complex are overrepresented for the Cellular Component domain. Amongst the top 20 differentially expressed genes (Table 5B, Table A2), are those encoding for the oxidoreductases plastocyanin and cytochrome c_{553} proteins (1\textsuperscript{st} and 3\textsuperscript{rd} most differentially expressed respectively) which have an important function in photosynthetic electron transport. In addition, psbA3 (Tery_4763) encoding for the Fe-binding PSII core protein D1 is the 15\textsuperscript{th} most differentially expressed gene.

What is more, the Molecular Function domain is enriched in genes coding for calcium-ion-binding-proteins which also appear to be heavily represented in the top 20 differentially expressed genes. The function of such proteins in cyanobacteria is currently not clearly characterised (Sec 5.3.8).

Other highly differentially expressed genes encode for the potentially Fe binding Tery_4114 (Mo-dependent nitrogenase-like) and the heme binding nirA (nitrite reductase, Tery_1068) (14\textsuperscript{th} and 19\textsuperscript{th} differentially expressed genes respectively) with the former hypothesised- and the latter known- to be involved in Trichodesmium’s nitrogen metabolism.

Also worth mentioning, Tery_1687, encoding for the glycolytic enzyme fructose-bisphosphate aldolase class I, is the 20\textsuperscript{th} most differentially expressed gene. The protein was previously observed to substitute the Fe containing class II fructose-bisphosphate aldolase in diatoms (Allen et al. 2012; Lommer et al. 2012) and Trichodesmium (Snow et al. 2015) under Fe-stress.
Figure 5.15: GO enrichment in differentially expressed genes. Select enriched GO categories as identified through singular enrichment analysis (SEA- AGRIGO) classified within the domains of cellular component, molecular function and biological process are presented for the differentially expressed genes (yellow- red bars) compared to the background of all the identified genes included in the analysis (salmon bars).
A large number of *Trichodesmium* genes are not yet annotated with several annotated as ‘hypothetical proteins’. An example is members of the Tery_08xx operon (discussed later) which is heavily represented (5 genes) in the top 20 differentially expressed genes. The second most differentially expressed gene identified in the experiment is also annotated as a hypothetical protein. This gene was previously identified as one of two *phoX*1 alkaline phosphatase homologues in *Trichodesmium* (Orchard et al. 2009). Interestingly, it was recently reported that PhoX requires both calcium and Fe for its production (Yong et al. 2014) and that its activity is influenced by phosphate but also desert dust and zinc (Mahaffey et al. 2014).

**Table 5B:** The top 20 differentially expressed genes presented is descending order of statistical significance (P-value).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Annotation</th>
<th>Cluster</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tery_2561</td>
<td>cytochrome c, class I (<em>petJ</em>)</td>
<td>2</td>
<td>1.09E-118</td>
</tr>
<tr>
<td>Tery_3845</td>
<td>hypothetical protein (<em>phoX1</em>)</td>
<td>6</td>
<td>1.85E-51</td>
</tr>
<tr>
<td>Tery_2563</td>
<td>plastocyanin</td>
<td>10</td>
<td>1.76E-48</td>
</tr>
<tr>
<td>Tery_0419</td>
<td>hemolysin-type calcium-binding region</td>
<td>2</td>
<td>1.25E-39</td>
</tr>
<tr>
<td>Tery_2055</td>
<td>hemolysin-type calcium-binding region</td>
<td>10</td>
<td>4.28E-38</td>
</tr>
<tr>
<td>Tery_0424</td>
<td>hemolysin-type calcium-binding region</td>
<td>2</td>
<td>4.91E-37</td>
</tr>
<tr>
<td>Tery_0850</td>
<td>hypothetical protein</td>
<td>2</td>
<td>1.00E-33</td>
</tr>
<tr>
<td>Tery_3467</td>
<td>hemolysin-type calcium-binding region</td>
<td>6</td>
<td>2.54E-28</td>
</tr>
<tr>
<td>Tery_1500</td>
<td>hypothetical protein</td>
<td>6</td>
<td>2.14E-25</td>
</tr>
<tr>
<td>Tery_0845</td>
<td>TENA/THI-4 protein</td>
<td>2</td>
<td>1.14E-22</td>
</tr>
<tr>
<td>Tery_3659</td>
<td>hypothetical protein</td>
<td>5</td>
<td>5.00E-22</td>
</tr>
<tr>
<td>Tery_3376</td>
<td>phosphoglycerate kinase (<em>pgk</em>)</td>
<td>5</td>
<td>1.35E-21</td>
</tr>
<tr>
<td>Tery_0846</td>
<td>hypothetical protein</td>
<td>2</td>
<td>2.48E-21</td>
</tr>
<tr>
<td>Tery_4114</td>
<td>Mo-dependent nitrogenase-like</td>
<td>10</td>
<td>5.20E-21</td>
</tr>
<tr>
<td>Tery_4763</td>
<td>photosystem q(b) protein</td>
<td>4</td>
<td>8.77E-21</td>
</tr>
<tr>
<td>Tery_0849</td>
<td>cell surface protein</td>
<td>2</td>
<td>1.96E-20</td>
</tr>
<tr>
<td>Tery_2976</td>
<td>caffeoyl-CoA O-methyltransferase</td>
<td>2</td>
<td>1.96E-20</td>
</tr>
<tr>
<td>Tery_0847</td>
<td>5-methyltetrahydropteroylglutamate-thonocysteine methyltransferase (<em>metE</em>)</td>
<td>2</td>
<td>1.11E-19</td>
</tr>
<tr>
<td>Tery_1068</td>
<td>ferredoxin-nitrite reductase (<em>nirA</em>)</td>
<td>7</td>
<td>6.28E-19</td>
</tr>
<tr>
<td>Tery_1687</td>
<td>fructose-bisphosphate aldolase</td>
<td>4</td>
<td>1.57E-18</td>
</tr>
</tbody>
</table>
5.3.6 Responses to FeCl₃ and dust-additions- Pairwise comparisons

Varied expression profiles across the four treatments Fe+, Fe-, Dust+ and [Dust] were previously recognised (Fig. 5.13). Four main drivers shaping these include:

A. **Iron**: Both in the form of FeCl₃-EDTA (Fe+) and Sahara Desert dust (Dust+)

B. **FeCl₃-EDTA (FeCl₃)**: Response stimulated either due to higher concentrations of Fe provided overall in Fe+, or due to increased levels of Fe’ provided compared to particulate/colloidal dust-bound Fe.

C. **Dust**: Induced through contact of the cells with dust (Dust+)

D. **Other chemical stimuli** leached from Dust: Independently or in combination with Fe

To better describe the changes induced by iron, dust and FeCl₃ the similarities and differences between the pairwise comparisons: Fe+ vs Fe- and Dust+ vs [Dust] are examined. We identify 376 genes regulated by dust and 438 genes by FeCl₃ from which 138 are coregulated by both drivers.

The response to contact with dust is biased towards upregulation in all of all three GO domains of Cellular Component, Molecular Function and Biological Process (Fig. 5.16). In contrast, FeCl₃ causes a more balanced response of upregulation: downregulation as do other stimuli from dust (comparison of [Dust] and Dust+ against Fe+ and Fe-). GO enrichment analysis for genes uniquely changed by contact with dust, FeCl₃ or co-influenced by both, indicated that dust alone has an effect on translation-related genes (Fig. 5.17). This possibly suggests that dust-related stimuli (only effective upon contact of cells with dust) trigger protein production in *Trichodesmium* and an increase of metabolic activity. The expected increase in transcription in such a scenario (to facilitate faster translation) would explain the large bias towards upregulation compared to downregulation of dust-affected genes.

It has previously been noted that the process of translation is downregulated as a response to stress related conditions (López-Maury et al. 2008; Singh et al. 2010) in order for energy reserves to be directed towards cell protection instead. Here, having an abundance of nutrients provided by dust might act as a trigger to boost *Trichodesmium* activity. Nevertheless, we cannot exclude a possibility that the more slow growing (compared to Fe+) Dust+ cultures are sampled at an earlier exponential phase when protein production is increased.
Iron in general (both as FeCl₃ and dust) induced processes such as photosynthesis, ATP synthesis, electron transport, phycobilisome proteins and Fe binding proteins. This is expected due to the large Fe demand of photosynthetic processes (Raven et al. 1999; Shi et al. 2007). At the same time, a different suit of phycobilisome proteins is upregulated only in response to dust while FeCl₃ alone is identified to significantly stimulate the light reaction of photosynthesis. The latter, involves the Fe demanding PSI and PSII requiring 12 and 3 atoms respectively (Shi et al. 2007), and includes the process of cyclic flow of electrons around PSI (which does not involve PSII). Protein components involved in cyclic electron transport were previously identified to be affected by Fe-limitation at the proteomic level in *Trichodesmium* (Snow et al. 2015). Changes in expression of genes involved in photosynthetic processes are reviewed in more detail later in this report.
Figure 5.17: Comparisons of transcriptomic changes due to additions of Dust and FeCl$_3$. (A). The number of genes affected in the Fe+ (yellow) and Dust+ (brown) treatments, with Fe- and [Dust] as background controls respectively, are presented in relation to each other to reveal genes regulated by both Fe sources (Bi-iii) Singular enrichment analysis of GO annotated genes is performed for the FeCl$_3$ (i), Dust (ii) and commonly Fe regulated genes (iii).
5.3.7 Markers of iron limitation. Confirmation of the cellular Fe state

From a set of 12 genes known to be regulated by Fe and previously used as markers to indicate its limitation, we identify that 5 are shaped by Fe both when provided as FeCl₃ and dust (fld2, petE, petI, fur1 and fur2) while an additional 3 are controlled solely by FeCl₃ (fbaB, isiA and isiB) (Table 5C).

The ferric uptake regulator genes 1 and 2 (fur1, fur2) (but not fur3) are differentially expressed in response to both Fe sources, with upregulation in Fe- and [Dust] compared to Fe+ and Dust+. Recent data demonstrate Fur’s role as both a repressor and activator of genes involved in Fe metabolism and regulator of the cellular redox status (González, M Teresa Bes, et al. 2012; González et al. 2016).

The 5 genes regulated by both Fe sources also include the electron transporters cytochrome c₅₅₃ (petJ), binding heme, and plastocyanin (petE), binding copper, mentioned previously to be the first and third most affected genes respectively. These interchangeable oxidoreductases are known to engage in a behaviour referred to as the plastocyanin-cytochrome switch regulated by copper and Fe concentrations (Wood 1978; De la Cerda et al. 2007). Here, in agreement to their characterised behaviour in other microbes, expression of petJ is increased while that of petE is reduced in Fe+ and Dust+ compared to Fe- and [Dust] treatments.

Another electron transporter that is switched for an alternative, Fe-cheaper protein, is ferredoxin. Under Fe limiting conditions this is known to be replaced by the flavin-containing, flavodoxin (LaRoche et al. 1996). Different expression patterns are observed here for the two Trichodesmium homologues of flavodoxin, with fld2 upregulated in Fe- and [Dust] relative to Fe+ and Dust+, while expression of fld1 (isiB) is upregulated in all conditions except Fe+. The latter is colocalised with the, similarly expressed, isiA gene in the Trichodesmium genome. IsiA proteins form an antennae surrounding the Fe demanding PSI complexes and thus enable a higher light trapping efficiency by fewer PSI units as a way of conserving Fe resources (Bibby et al. 2001). IsiA and IsiB proteins were also identified to differ at the protein level in a previous similar study (Snow et al. 2015).
Table 5C: Expression of suggested biomarkers of iron limitation. The regulation of genes by FeCl₃ only (yellow) or in combination with dust (orange) is compared to previously identified differential regulation at the protein level (blue, Snow et al 2015). No dust only changes were observed in this group of genes.

<table>
<thead>
<tr>
<th>Category</th>
<th>gene</th>
<th>Tery number</th>
<th>RNA changes</th>
<th>Protein changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photosynthesis</td>
<td>Iron stress induced protein A (IsiA)</td>
<td>1667</td>
<td>Yellow</td>
<td>Blue</td>
</tr>
<tr>
<td></td>
<td>Flavodoxin (isiB/fld1)</td>
<td>1666</td>
<td>Yellow</td>
<td>Blue</td>
</tr>
<tr>
<td></td>
<td>Flavodoxin (fld2)</td>
<td>2559</td>
<td>Orange</td>
<td>Orange</td>
</tr>
<tr>
<td></td>
<td>Plastocyanin</td>
<td>2563</td>
<td>Orange</td>
<td>Orange</td>
</tr>
<tr>
<td></td>
<td>Cytochrome c₅₅₃</td>
<td>2561</td>
<td>Orange</td>
<td>Orange</td>
</tr>
<tr>
<td>Storage</td>
<td>Bacterioferritin</td>
<td>2787</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ferritin</td>
<td>4282</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulation</td>
<td>Ferric uptake regulator1</td>
<td>1958</td>
<td>Orange</td>
<td>Orange</td>
</tr>
<tr>
<td></td>
<td>Ferric uptake regulator2</td>
<td>3404</td>
<td>Orange</td>
<td>Orange</td>
</tr>
<tr>
<td></td>
<td>Ferric uptake regulator3</td>
<td>1953</td>
<td>Orange</td>
<td>Orange</td>
</tr>
<tr>
<td>Nitrogen fixation</td>
<td>NifH</td>
<td>4136</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycolysis</td>
<td>Fructose bisphosphate aldolase (Class I)</td>
<td>1687</td>
<td>Yellow</td>
<td>Blue</td>
</tr>
</tbody>
</table>
Similarly to *isiB* and *isiA*, expression of the class I fructose-bisphosphate aldolase gene (*fbaB*) is suppressed only by additions of FeCl$_3$ and not dust. This protein is switched with class II fructose bisphosphate aldolase when Fe is abundant (Snow et al. 2015), a change not identified here at the transcriptomic level.

Regulation by FeCl$_3$ but not dust indicates that some genes considered as good Fe-limitation biomarkers (*isiA*, *isiB* and *fbaB*) are either not sensitive to particulate/colloidal Fe or that the concentration acquired from the latter is not sufficient to suppress the expression of these stress related genes. This is the first time the transcriptomic response of these genes to dust is assessed in culture and indicates that caution should be applied before their use as markers in the field.

5.3.8 Transcriptomic response of major processes across the experimental treatments.

The effect of desert dust and FeCl$_3$ to processes expected, or previously indicated to be affected by Fe-limitation, including Fe transport, photosynthesis, energy production and nitrogen metabolism is discussed below. In addition, putative outer membrane components potentially involved in the Fe-limitation response of *Trichodesmium* are identified.

5.3.8.1 Iron transport

Iron is carried across the cyanobacterial membranes through various pathways depending on its speciation and complexation. Expression of the *Trichodesmium* genes involved in these pathways, as inferred by homology, is reviewed below.

The ferrous iron uptake system Feo, first characterised in *E.coli*, is composed of three proteins (Kammler et al. 1993): FeoA which is suggested to interact with a conserved transmembrane domain of the inner membrane ATP powered transporter, FeoB, and FeoC a transcriptional regulator (Cartron et al. 2006; Lau et al. 2013). In the model cyanobacterium *Synechocystis feoB* is upregulated under Fe-stress while its mutant shows reduced Fe$^{2+}$ uptake compared to the WT (Katoh et al. 2001b; Kranzler et al. 2014). Kranzler et al. (2014) suggest that the FeoAB is the major Fe transport system in this organism as minimal uptake is observed in the Δ*feoB* mutant. *Trichodesmium* possess *feoAB* homologues and *feoB* transcription was shown to be upregulated under Fe-limitation (Chappell & Webb 2010). This observation is supported in the present transcriptome as expression in FeCl$_3$ added cultures is lower than that of the other treatments although the
difference is not statistically significant (Fig. 5.19). Contrariwise, feoA expression is low compared to feoB and its level is stable across all treatments.

Oxidised Fe’ (Fe$^{3+}$) is transferred across the inner membrane using the ferric ion binding protein (Fbp) system first described in pathogenic bacteria (Krewulak & Vogel 2008). It is composed of a periplasmic iron-binding component (FbpA), a transmembrane protein (FbpB), and an ATPase (FbpC) subunit which in *Synechocystis* are referred to as FutABC. Deletion of *futBC* in this species causes reduced growth (Katoh et al. 2001a), while deletion of the two *futA* homologues interferes with Fe uptake and reduction (Kranzler et al. 2014). Further information regarding *Synechocystis futA1* and *futA2* is provided in Chapter 4. The expression of the *Trichodesmium futA* but not *futBC* is significantly lower in the presence of FeCl$_3$ compared to other treatments (Fig. 5.18, 5.19).

Figure 5.18: Iron uptake across the inner and outer cell membranes. Representation of *Trichodesmium* putative Fe transporters and related proteins as identified by homology to studied pathways of other organisms. Shading indicates changes at the RNA level in response to FeCl$_3$ both uniquely (yellow) and in combination with dust (orange). No changes specific to contact with dust are identified in the expression of genes for this set of proteins. Differences between the Dust+ and [Dust] cultures relative to Fe- and Fe+ (green) potentially indicate changes in transcription stimulated by dust-released chemicals.

Both the abovementioned uptake systems are involved in transfer of Fe’ through the cyanobacterial inner membrane. Although small molecules are able to pass through porins of the outer membrane (Nikaido 2003), molecules bigger than 600 Da are actively transported to the periplasmic space by TonB- dependent transporters (TBDTs). TBDTs involved in nickel and cobalt (Schauer et al. 2008), disaccharide and polysaccharide as well as vitamin B12 uptake have been identified, but siderophores and heme are considered their major substrates (Hopkinson & Barbeau 2012). These transporters are powered by
energy transferred through the periplasm-spanning TonB anchored on the plasma membrane. For this energy transduction the auxiliary proteins ExbB and ExbD located in the plasma membrane are also required (Ollis et al. 2009). Recent evidence suggest that in addition to siderophore uptake Synechocystis ExbBD are also involved in inorganic Fe uptake (Jiang et al. 2015).

Under Fe starvation, differential expression of tonB and exbBD has been reported for the model diazotroph Anabaena sp. PCC 7120 (Anabaena) (Stevanovic et al. 2012) and of exbBD for Synechocystis (Singh et al. 2003; Jiang et al. 2015). Under the same conditions, mutants of tonB and exbBD in Synechocystis display reduced growth compared to the WT. In addition, the four Synechocystis TBDTs are upregulation under Fe-limitation (Katoh et al. 2001a; Jiang et al. 2012). Anabaena has 22 TBDTs which have been shown to be differentially expressed under variable copper/Fe and nitrogen concentrations (Mirus et al. 2009; Stevanovic et al. 2013). The function of only some of these TBDTs has been researched, including SchT involved in schizokinen/aerobactin uptake, IutA2 involved in schizokinen uptake as well as IacT which takes up Fe and copper (Cu) (Nicolaisen et al. 2008; Nicolaisen et al. 2010; Rudolf et al. 2016).

Trichodesmium on the other hand, possesses homologues of ExbBD and TonB but does not appear to have any obvious TBDTs (Hopkinson & Morel 2009). Expression of exbD but not exbB is upregulated in the Fe+ and Fe- treatments compared to the [Dust] and Dust treatments here. The two tonB homologues also show a similar expression profile but only tonB2 is significantly increased in Fe- and Fe+ compared to [Dust] and Dust+ (Fig. 5.18, 5.19). The results suggest that in Trichodesmium this transport system might be associated with a soluble substrate released from the desert dust.

Adjacent to tonB1 but transcribed from the opposite DNA strand, Tery_1561 annotated as a hypothetical protein, has suppressed expression in response to Fe as both FeCl3 and dust. compared to the Fe-deficient Fe- and [Dust] treatments. The protein appears to contain an alpha/beta hydrolase fold (NCBI Conserved Domain Database, Marchler-Bauer et al. 2015) which if linked to Fe uptake could be speculated to contribute to extraction of Fe from ligands prior to uptake through TBDTs. In addition, the adjacent Tery_1562 is homologous to a copper translocating P-type ATPase. This is more highly expressed in Fe- compared to Fe+ and it remains at low levels in both the Dust+ and [Dust] treatments. A possible association of this proteins to TonB and any involvement to Fe transport remains to be examined.
Furthermore, transcription of heme oxygenase (Tery_0335), catalysing the degradation of heme, is decreased in Fe- compared to Fe+ as well as the dust treatments Dust+ and [Dust] although the former is not statistically different. The increased expression in Fe+ compared to Fe- is also shown at the protein level (Snow et al. 2015). Additionally, the same study identifies proteins Tery_3823-3826 as potentially involved in siderophore production and uptake and indicates differential expression between Fe+/Fe-. Here, increased expression at the transcriptomic level is identified in Fe+ and Dust+ compared to Fe- and [Dust] for Tery_3824, and in Fe+ compared to Fe- for Tery_3825-3826 reinforcing the author’s suggestions.

Once inside the periplasm, siderophores that are taken up by bacterial TBDTs can be carried across the inner membrane through specialised ABC transporters. The Fhu system, best characterised in proteobacteria, involves binding of hydroxamate siderophores by the periplasmic FhuD protein and their transport across the inner membrane FhuB. The process is powered by the ATP binding FhuC. In a similar way, citrate-type siderophores are transferred by the Fec system (Krewulak & Vogel 2008).

In Anabaena, which possesses both of the abovementioned systems, fhu genes are upregulated under Fe-deficient conditions and in mutants of the siderophore secretion and uptake pathways. Also, deletions of the Fhu uptake system components cause reduced growth compared to the WT under Fe-limitation (Stevanovic et al. 2012). Furthermore, three Fec systems have been identified so far in this organism (Nicolaisen et al. 2008; Mirus et al. 2009; Stevanovic et al. 2012) and components were shown to be affected by Fe, copper and nitrogen concentrations (Stevanovic et al. 2012; Stevanovic et al. 2013). In Synechocystis members of the Fec system were also identified but there was no obvious upregulation under Fe-limitation (Katoh et al. 2001a).

Contrary to Anabaena and Synechocystis, Trichodesmium does not appear to possess members of these two pathways except from a homologue to the periplasmic hydroxamate binding FhuD (Chappell & Webb 2010). This could be either due to the existence of a distinct, yet to be discovered, siderophore uptake transport pathway in this organism, or the dissociation of Fe from ligands prior to uptake through the inner membrane. In the presented analysis expression of fhuD is increased in Fe- and [Dust] compared to Fe+ and Dust+ conditions (only Fe- compared to Fe+ are significantly different) (Fig. 5.18, 5.19).
Independently or in synergy to the previously described systems, Fe can be reduced either before or after its transfer through the outer membrane. Reduction pathways in eukaryotic phytoplankton include the use of cell surface ferric reductase (FRE) similar to that of Saccharomyces cerevisiae (Yun et al. 2001; Kustka et al. 2007; Groussman et al. 2015). A ferric-reductase-like gene (Tery_2227) identified in the Trichodesmium genome shows no differential expression under any condition (Fig.5.18).

Moreover, in yeast and eukaryotic phytoplankton like T.pseudonana and C.reinhardtii, after reduction by FRE and consequent release from organic ligands, Fe can be reoxidised to Fe$^{3+}$ by multicopper oxidases (MCO) to be transported through Fe$^{3+}$ specific transporters. An MCO with Fe$^{3+}$ oxidising function has not yet been characterised in cyanobacteria. Here, a gene homologue to multicopper oxidases (Tery_0359) in Trichodesmium demonstrates increased expression in Fe- compared to Fe+ (Fig. 5.18, Fig. 5.19). However, prediction software (Signal P, Petersen et al. 2011) does not identify a likely signal sequence to this protein, either indicating an intracellular function or a divergent sequence for export from the cytoplasm.

Ferric (Fe$^{3+}$) transporters associated with MCO include permeases (Ftr1) (Stearman et al. 1996; Askwith & Kaplan 1998; Urbanowski & Piper 1999; Wang & Dei 2003; Bonaccorsi di Patti et al. 2005) and transferrins (TTf) (Fisher et al. 1997; Fisher et al. 1998; Paz et al. 2007b). No homologues of such proteins are identified in the Trichodesmium genome. However, immediately downstream of multicopper oxidase- like Tery_0359, Tery_0360 is indicated as a transmembrane protein (contains the Tmemb_14 domain of transmembrane uncharacterised proteins). Similarly to Tery_0359, this gene is upregulated in Fe- compared to Fe+. Further evidence is required to identify the function of this protein and whether in association with the putative MCO they constitute a Trichodesmium Fe uptake system.

Although similarly to Lyngbya majuscula (Rose & Waite 2005) the use of electron donating reactive oxygen species (ROS) in the form of superoxide has been suggested for Trichodesmium (Roe & Barbeau 2014; Hansel et al. 2016), the pathway of this process is not characterised and thus could not be investigated using the molecular data collected in the context of this work.

A different hypothesised mechanism for reduction involves transfer of electrons to Fe$^{3+}$ through the plasma membrane-located alternative respiratory terminal oxidase (ARTO). Evidence for this was the observation that the Synechocystis ARTO mutant displays.
Figure 5.19: Gene expression of various components of putative Fe uptake pathways in *Trichodesmium*. Expression of genes possibly involved in transport of Fe either as (A) Fe$^{3+}$ or (B) Fe$^{2+}$ across the inner cell membrane, (C) Fe$^{2+}$ oxidation prior to uptake and (D-F) siderophore uptake, is presented across the experimental treatments [Dust] (light brown), Dust + (dark brown), Fe- (green) and Fe+ (yellow). Error bars represent standard deviation of the means of three biological replicates.
impaired Fe reduction and uptake (Kranzler et al. 2014). Possibly suggesting a similar function in *Trichodesmium*, the ARTO genes (*Tery_0276-0278*) are upregulated in Fe- and [Dust] compared to Fe+ and Dust+ (Fig. 5.18).

5.3.8.2 Photosynthetic Electron Transport Chain

Genes involved in the process of photosynthesis are amongst the most highly expressed (Table A3) and enrichment analysis suggests that it is one of the most heavily affected processes across the experimental treatments, with the response of different components depending on the form of Fe provided (Fig. 5.17). Although Fe as both dust and FeCl$_3$ have common effects on a group of photosynthesis related genes, dust additions exclusively cause differential expression in thylakoid and phycobilisome genes while FeCl$_3$ causes GO enrichment in the categories of the photosynthetic light reaction and electron transport. Below the responses of genes encoding for each component of the photosynthetic electron transport chain are reviewed separately.

**PSII:** This protein-pigment complex, comprised of around 30 proteins, utilises solar energy to oxidise water and reduce plastoquinone (PQ) starting a chain of electron transfer that will ultimately result in reduction of NADP to NADPH (Vinyard et al. 2013). Subunits cytochrome $c_{550}$ and $c_{559}$ bind heme and the D1/D2 core proteins bind non-heme Fe, resulting in a requirement of 3 Fe atoms per PSII monomer (Shi et al. 2007).

Genes *psbV* (*psbV1* and *psbV2*) and *psbE/psbF* for the heme containing cytochrome $c_{550}$ and $b_{559}$ respectively are upregulated in Fe+ and Dust+ compared to Fe- and [Dust] (Fig. 5.20). For the cytochrome $b_{559}$ beta subunit gene *psbF* these differences are not statistically significant. In contrast to the cytochrome subunits, genes *psbA1-3* and *psbD* for the non-heme Fe-binding D1/D2 are downregulated in Fe+ compared to Fe- (Fig. 5.20) and their levels remain high in the Dust+ and [Dust] treatments. Contradicting the expression pattern of *psbD* observed here, upregulation in Fe+ compared to Fe- was previously reported (Shi et al. 2007). At the proteomic level PsbD levels were suggested to remain stable in comparable conditions (Snow et al. 2015).

Genes *psbB* and *psbC1* forming intrinsic transmembrane CP43 and CP47 antennae proteins of PSII are upregulated in Dust+ compared to [Dust] while they remain at lower levels in both Fe+ and Fe (Fig. 5.20). Homologs *psbC2* and *psbC3* of the CP47 gene have reduced transcription in Fe+ relative to Fe- (Fig. 5.20) while the Dust+ and [Dust] expression levels remain high. Upregulation in Fe- compared to Fe+ is also observed for the *psbW* (*psb28*)
gene (Fig. 5.20), forming a subunit of the oxygen evolving complex with functions in biogenesis, assembly and stabilisation of PSII (Garcia-Cerdán et al. 2011).

Figure 5.20: Changes in transcription of genes involved in the process of photosynthesis. The differential expression of genes from pairwise comparisons between (A) Fe- to Fe+ and (B) [Dust] to Dust+ involved in electron transport (green) or encoding for subunits of F\textsubscript{}\text{O}_\text{O}_\text{F} ATPase (blue), NADPH dehydrogenase (orange), PSI (black) and PSII (purple) is plotted in log scale. The line y=x separates upregulated (below) from downregulated (above) of genes upon additions of FeCl\textsubscript{3} (A) and Dust (B).
Results also suggest that possible release of dust-leached chemicals influences transcription of PSII subunit genes. This is indicated by an upregulation of PSII genes for accessory subunits PsbHIJKLNOTUZ27 and one of the PsbA isoforms, in Dust+ and [Dust] compared to the Fe- and Fe+ treatments.

**Cytochrome b_{6f}:** This thylakoid membrane located enzyme catalyses the transfer of electrons from plastoquinol to plastocyanin/cytochrome c_{553} and generates a proton gradient by pumping H^{+} through the thylakoid membranes into the lumen (Kurisu et al. 2003).

The cytochrome b_{6f} complex is made of 4 core subunits two of which, PetB (2 Fe, cytochrome b_{6}) and PetD (subunit IV), are membrane intrinsic, and two, PetA (1 Fe, cytochrome f) and PetC (2 Fe, Rieske Fe-S protein), are membrane extrinsic (Kurisu et al. 2003). The genes for each set are colocalised in the *Trichodesmium* genome.

The two clusters appear to be under distinct regulatory control. Genes petB and petD are upregulated under Fe- compared to Fe+ conditions (statistical significance only for petD) while genes petA and petC are more highly expressed in the dust treatments, Dust+ and [Dust], compared to Fe- and Fe+. Gene petG potentially involved in assembly and stability of cytochrome b_{6f} (Schneider et al. 2007) is also upregulated in Dust+ and [Dust] treatments.

Our results contradict those of previous studies reporting a downregulation of cytochrome b_{6f} in Fe- conditions in *Trichodesmium* (Shi et al. 2007; Snow et al. 2015), the diatom *T.oceanica* (Lommer et al. 2012) and *Prochlorococcus* (Thompson et al. 2011) indicating possible dependency on the growth phase or other unknown factors.

**PSI:** The second light- absorbing protein complex of photosynthesis, PSI, mediates transfer of electrons from plastocyanin/cytochrome c_{553} to ferredoxin which will in turn reduce ferredoxin-NADP+ reductase (FNR) the protein generating NADPH from NADP (Jordan et al. 2001). PSI is composed of multiple subunits (around 15) and is more Fe demanding compared to PSII, binding 12 Fe atoms per monomer (Shi et al. 2007). These are associated with the 4Fe-4S iron-sulphur centres F_{X}, F_{A} and F_{B} (Jordan et al. 2001).

The PSI PsA and PsB subunits form heterodimers that bind the 4Fe-4S iron-sulphur centre F_{X}. They also interact with the electron donor P700 (primary electron donor of PSI) and the acceptors A_{0}, A_{1} (Jordan et al. 2001). Genes psaA and psaB are downregulated in Fe- and
[Dust] conditions compared to Fe+ and Dust+ (Fig. 5.20) as also observed at the protein level between Fe- and Fe+ (Snow et al. 2015).

The F_A and F_B 4Fe-4S centres are bound to PsaC (Jordan et al. 2001), previously identified to be upregulated in response to FeCl_3 (Snow et al. 2015). Gene psaC is not differentially expressed here under similar conditions, and although expression is lower in [Dust] compared to Dust+ this change is not statistically significant. However, similarly to genes psalJ (encoding for subunits Psal and PsaJ), a secondary controlling factor stimulates increased expression of psaC in the [Dust] and Dust+ treatments compared to Fe- and Fe+.

Furthermore, genes for PsaD, which can form complexes with the electron transporter ferredoxin, and PsaF the plastocyanin-docking protein important in electron transfer from plastocyanin to P700 (Jordan et al. 2001) are upregulated in Dust+ compared to [Dust] (Fig. 5.20).

**PSI/PSII expression comparison:** It is interesting to note that the PSI or PSII genes affected by FeCl_3 have an opposite response to each other. PSII genes (apart from the heme containing psbE and psbV) are upregulated in Fe- compared to Fe+ while the PSI genes that are significantly changed (psaAB) are downregulated (Fig 5.20, 5.21). It is possible that this reflects the shift of PSI: PSII ratio to reduce the more Fe demanding PSI compared to PSII under Fe deficient conditions (Snow et al. 2015). The plasticity of the PSI: PSII ratio was also previously documented in comparisons between coastal and open ocean phytoplankton growing under Fe-deficiency (Strzepek & Harrison 2004).

However, the lower PSI:PSII potentially negatively impacts cyclic electron transport (CET) (Berman-Frank et al. 2001; Snow et al. 2015). Through CET, NADPH generated by FNR can be reoxidised and the electrons released, recirculated through cytochrome b,f and PSI to fuel additional ATP production without the accumulation of NADPH normally occurring during linear electron transport (LET) (Lea-Smith et al. 2016). Also, the reduction of CET can potentially lead to increased oxygen evolution and ROS production and is expected to affect N_2 fixation through oxygen deactivation of its key enzyme, nitrogenase (Milligan et al. 2007).

**Electron transport:** The petE and petU genes encoding for the electron transporters plastocyanin and cytochrome c_553 respectively are regulated inversely to each other both in response to FeCl_3 and dust as the Fe containing cytochrome c_553 is replaced by plastocyanin in Fe-deficient conditions (Fig. 5.20, 5.21) (Wood 1978; De la Cerda et al. 2007).
Another known Fe-stress response involves switching of the Fe containing electron transporter ferredoxin to flavodoxin (La Roche et al. 1996). The non-Fe containing flavodoxin genes \(fld1\) and \(fld2\) are upregulated in Fe- compared to Fe+. Gene expression of \(fld2\) is also stimulated in [Dust] compared to Dust+ (Fig. 5.20, 5.21). However, from the 5 ferredoxin encoding \(petF\) homologues found in the \(Trichodesmium\) genome, 2 are upregulated in Fe- compared to Fe+ treatments (Fig. 5.20, 5.21) and are kept at high levels in Dust+ and [Dust] while the remaining 3 are not significantly changed.

Finally, the ferredoxin NADP+ reductase (FNR) encoding, \(petH\), is upregulated in Dust+ and Fe+ compared to [Dust] and Fe- treatments with the difference being significant only in response to dust (Fig. 5.20, 5.21).

\(F_{1}F_{o}\) ATP synthase: The proton gradient generated from photolysis of water by PSII and shuttling of H+ by cytochrome \(b_{6}f\) stimulates \(H^{+}\) transport down the electrochemical gradient through the \(F_{o}\) proton porin of the ATP synthase. The \(F_{1}\) motor consequently catalyses the formation of ATP from ADP and Pi (Berg et al. 2002).

Components alpha, gamma, delta of \(F_{1}\) and b of \(F_{o}\) are upregulated in Fe+ and Dust+ compared to Fe- and [Dust] treatments whereas subunits a and c of \(F_{o}\) are only upregulated in response to dust but not \(FeCl_{3}\) (Fig. 5.20, 5.21).

\(NAD(P)H\) dehydrogenase: \(NAD(P)H\) dehydrogenase I (Ndh-1), is an enzyme composed of around 15 subunits which can be involved in PSI CET as well as respiration and \(CO_{2}\) uptake (Ohkawa et al. 2000b; Battchikova et al. 2011; Lea-Smith et al. 2016). Although most subunits are shared between Ndh-1 performing different functions, NdhD and NdhF seem to specialise according to the activity: Gene deletions in \(Synechocystis\) reveal involvement of \(ndhD1\) in PSI CET and \(ndh3\) in \(CO_{2}\) uptake (Ohkawa et al. 2000a; 2000b; Shibata et al. 2001; 2002; Maeda et al. 2002; Battchikova et al. 2011).

\(Trichodesmium\) has 3 \(ndhD\) paralogues with homology to the \(Synechocystis ndhD1, ndhD2\) and \(ndh3\) genes Gene \(ndhD2\) is stimulated in Fe- compared to Fe+ while \(ndhD1\) in Dust+ compared to [Dust] (Fig. 5.20). Upregulation of \(ndhD2\) in Fe- conditions (as well as P, N, \(CO_{2}\) and S-limitation and high light) was also previously identified in \(Synechococcus\) PCC 7002 (Ludwig & Bryant 2011; Ludwig & Bryant 2012). The function was associated with balancing the ratio of protons translocated across the membrane to that of electrons in Fe- conditions when the availability of reducing equivalents is lower.
Figure 5.21: Transcriptomic response of subunits involved in the photosynthetic electron transport chain. Positive (green) or negative (red) changes caused by (A) FeCl₃ and (B) Dust-additions (relative to Fe- and [Dust] respectively) are compared. Positive/negative changes indicated are based on the direction of changes of the majority of genes for each subunit. The number of upregulated genes/downregulated genes is presented for each component (bottom right).
Similarly to ndhD1, genes ndhH and ndhF1 are upregulated in Dust+ compared to [Dust] (Fig. 5.20). The former is also more highly expressed in both of the Dust+ and [Dust] treatments compared to Fe- and Fe+.

Subunit NdhI which is responsible for the transfer of electrons from an electron donor to NdhK and eventually plastoquinone, harbours two iron-sulphur clusters (N6a, N6b) (Battchikova et al. 2011). In the presented transcriptome gene ndhI is upregulated in Fe- and [Dust] compared to Fe+ and Dust+.

Light-harvesting antennae: The phycobilisome antennae made of the phycobiliproteins allophycocyanin (core), phycocyanin/phycoerythrocyanin (linker) and phycoerythrin (rods) are known to attach and transfer light energy mainly to PSII but also to PSI. Their synthesis has high Fe requirements (Grossman et al. 1993). In the presented transcriptome, genes for most of these components are upregulated in Dust+ and Fe+ compared to Fe- and [Dust]. This was also previously observed for Synechococcus sp. PCC 7942 and Synechocystis (Sandström et al. 2002; Schrader et al. 2011).

Although the structure of the PSII phycobilisome antennae has been well studied the PSI antenna is less well understood. Recent evidence indicates that subunit CpcG3 of Anabaena sp. PCC 7120 (Watanabe et al. 2014) and CpcG2 of Synechococcus PCC 7002 (Deng et al. 2012) have an important function in the organisation of the PSI phycobilisome antenna. Upregulation of cpcG2 in [Dust] and Fe-, opposite to the remaining phycobilisome genes, could be indicative of an increase in PSI antennae to enhance the light harvesting efficiency and ameliorate the impact of a reduction in PSI subunits.

Finally, similarly to what is observed for various PSI and PSII subunit genes, apcD, apcF, cpcA, cpeC and cpeE3 are upregulated in Dust+ and [Dust] compared to Fe- and Fe+ treatments.

An overview: Changes of gene expression are observed in all the major electron transport chain components (Fig. 5.20). Dust+ stimulates expression of the majority of genes in all components except ferredoxin and cytochrome b_{5f} whose transcript levels remain unaffected, and the electron transporters flavodoxin and plastocyanin which are expressed at reduced levels compared to the [Dust] treatment. Flavodoxin and plastocyanin are known to replace ferredoxin and cytochrome c_{553} under Fe-deficiency (Wood 1978; LaRoche et al. 1996) and this is also confirmed in the comparison between Fe+ and Fe-gene expression, as observed here (Fig. 5.20).
On the other hand, upregulation of genes (the majority of the genes whose expression is altered) encoding NAD(P) dehydrogenase, PSII and cytochrome \( b_{6f} \) is observed in Fe-compared to Fe+ (Fig. 5.20). It is possible that this is a cellular attempt to compensate for a reduction of the Fe-expensive PSI subunits and phycobilisome antennae under Fe-stress.

Another hypothesis is that the rapidly growing Fe+ cultures deplete reserves of a nutrient that by contrast is not limiting in Dust+ and therefore permit upregulation of PSII and NAD(P) dehydrogenase in the latter (compared to [Dust]). This could also justify the general upregulation in dust cultures, ([Dust] and Dust+) compared to Fe+ and Fe- observed across the photosynthetic components.

It is important to recognise that these hypotheses cannot be strongly supported with transcriptomic data as changes in one, or several genes, encoding for a protein complex are not indicative of its expression and assembly at the protein level. However, they can serve to signify potentially interesting targets for further analysis.

5.3.8.3 Oxidative phosphorylation

In cyanobacteria oxidative phosphorylation utilises both the thylakoid and plasma membranes while members of the photosynthetic and respiratory electron transport chain plastoquinone, plastocyanin/cytochrome \( c_{553} \), cytochrome \( b_{6f} \) and NADH dehydrogenase are shared (Lea-Smith et al. 2016). Additional components in the respiratory electron transport chain are cytochrome oxidase (COX) and alternative respiratory terminal oxidase (ARTO) which reduce oxygen to water and pump H\(^+\) across the thylakoid and cell membranes to power ATP production (through ATP synthase) (Hart et al. 2005; Lea-Smith et al. 2016).

Genes encoding for subunits of ARTO (ctaCDE) are upregulated in Fe- and [Dust] compared to Fe+ and Dust+ respectively. Also, genes Tery\_0279 and Tery\_0280 for transmembrane ‘hypothetical proteins’ upstream of the ARTO genes are expressed in a similar fashion. These were previously shown to be co-transcribed with ARTO in Anabaena but their function is not characterised (Valladares et al. 2003). The potential importance of ARTO in Fe reduction prior to its uptake (Kranzler et al. 2014) needs to also be considered.

Similarly, to the ARTO genes, cytochrome c oxidase assembly genes \( cox15 \) and \( cyoE \) are regulated by Fe but only differences between [Dust] and Dust+ are statistically significant. The COX genes are not differentially expressed in response to Fe.
Chapter 5: Dust as a source of iron to *Trichodesmium*

However, expression of *coxA*, whose protein binds heme α and a heme α₂-Cu₃ centre (Alge et al. 1994), as well as the ARTO *ctaCD*, are lower in the dust treatments [Dust] and Dust+ compared to Fe- and Fe+. This indicates an additional effect of putative dust derived chemicals to the process of oxidative phosphorylation.

### 5.3.8.4 Nitrogen metabolism

Nitrogen (N₂) fixation is negatively affected by photosynthetic oxygen evolution and therefore strategies have evolved to separate the two processes. In *Trichodesmium* the peak N₂ fixation activity occurs at midday and coincides with a depression of photosynthetic activity (Berman-Frank et al. 2001). Samples from our experiment were acquired at 5 hrs through the 12 hr light period.

Contrary to previous reports (Shi et al. 2007; Snow et al. 2015) the genes encoding for the nitrogenase structural proteins NifD, NifH and NifK (*nifDHK*) were not differentially expressed. However, gene *nifZ*, involved in the maturation of the Mo-Fe protein P-cluster site containing a [4Fe-4S] and *fdxH*, encoding for the ferredoxin protein of nitrogenase are upregulated in Fe- relative to Fe+. In addition, *nifB* (involved in the biosynthesis of the Mo-Fe cofactor of nitrogenase) and *cysE* (involved in biosynthesis of cystein required for Fe-S cluster construction) are upregulated in Dust+ compared to [Dust].

The nitrate/nitrite utilisation genes *nirA* (nitrite reductase), *narB* (nitrate reductase) and *napA* (nitrate permease) are upregulated in Fe- and [Dust] cultures compared to Fe+ and Dust+ indicating the potentially N limited state of the cells in the former. A similar response in the expression of genes involved in nitrate/nitrite reduction and transport was observed under Fe-deficiency of *Synechocystis* (Hernández-Prieto et al. 2012). It should be mentioned that nitrate and nitrite reductase enzymes both have [4Fe-4S] cluster cofactors and therefore require Fe for their production (Luque et al. 1993; Rubio et al. 2002; Jepsoni et al. 2004).

Furthermore, upregulation of gene *gdhA* for the GdhA subunit of glutamate dehydrogenase in Fe+ compared to Fe- could be linked to an increased production of fixed nitrogen in Fe+, which is consequently processed through this enzyme to generate glutamate. Finally, regulatory genes of nitrate assimilation, *ntcA* and *ntcB* are not differentially expressed in the experimental treatments.
5.3.8.5 Outer membrane proteins

Haemolysin-type calcium binding (HTCaB) region domain: Growth of *Trichodesmium* on FeCl₃ and dust stimulates gene expression changes for proteins with HTCaB domains. The GO analysis identifies calcium-ion binding as an enriched category and 4 such genes are in the top 10 most differentially expressed genes identified in the transcriptome.

Calcium (Ca) has an important structural role in bacteria as part of the lipopolysaccharide layer and cell wall (Smith 1995, pp.83–133) and is involved in the cellular signalling system. A close link between the Ca and ROS signalling systems (Gordeeva et al. 2003; Brookes et al. 2004; Yan et al. 2006; Singh & Mishra 2015) denotes that changes in the former can be caused as an effect to stimulated ROS production under stress conditions. Such stress conditions are assumed to be encountered here in the Fe deficient, Fe- and [Dust].

Calcium binding domains in bacteria include HTCaB and EF-hand domains. HTCaB occur in tandem repeats in proteins that can form a parallel β roll structure (Baumann et al. 1993; Lilie et al. 2000) and are exported from the cell to function as haemolysin, cyclolysin, leukotoxin and metallopeptidases (Duong et al. 1992; Boehm et al. 1990; Rose et al. 1995). Adhesive properties (Sánchez-Magraner et al. 2007) and roles in motility (Brahamsha & Haselkorn 1996; Hoiczyk & Baumeister 1997; Pitta et al. 1997) have also been demonstrated.

*Trichodesmium*’s HTCaB proteins have not been previously studied but the presented transcriptomic analysis suggests an important role. Ten out of the 15 HTCaB genes in the *Trichodesmium* genome are differentially expressed with variable profiles across the 4 experimental treatments (Fig. 5.22A). Three of these have characteristic Fe+ to Fe- and Dust+ to [Dust] differences (*Tery_0419, Tery_0424 and Tery_2055*) and bioinformatic analysis indicates that their sequence includes multiple CHRD domains (identified in chordin). These domains are expected to have an immunoglobulin-like β-barrel structure based on some similarity to superoxide dismutases, enzymes with protective function against oxidative stress, but their function is currently unknown (Hyvönen et al. 2003). Also annotated as HTCaB, and upregulated under Fe+, *Tery_3467* was previously identified as the alkaline phosphatase (APase) gene *phoA* (Orchard et al. 2003). Although the involvement of Fe in assembly and regulation of the APase PhoX has been previously mentioned (Yong et al. 2014, Mahaffey et al. 2014), PhoA is thought to require zinc and magnesium (Coleman 1992, Jakuba et al. 2008). Results presented here show Fe regulation of PhoA suggesting possible Fe binding properties.
Figure 5.22: Gene expression of putative outer membrane proteins. Expression of genes with haemolysin-type calcium binding domains (HTCaB) (A-i-x) and filamentous hemagglutinin outer membrane domains (FHOM) (i-v) in the Fe- (green), Fe+ (yellow), [Dust] (light brown) and Dust+ (dark brown) treatments. Error bars are the standard deviations from the calculated means of triplicate cultures.
Filamentous hemagglutinin outer membrane (FHOM) domain proteins with similarity to the filamentous hemagglutinin family (containing the FhaB domain) consists of 10 large proteins in *Trichodesmium*, 5 of which are differentially expressed (Fig. 5.22) in the presented transcriptome. In pathogenic microorganisms such as *Bordetella pertussis* filamentous hemagglutinin forms adhesive molecules produced as virulence factors (Melvin et al. 2015). FhaB is the pre-pro protein that will be processed during export from the cell to form a long rod-shaped outer membrane product. Its translocation utilises the plasma membrane SEC export system and the two-partner secretion (Tps) pathway with the assistance of FhaC in the outer membrane. It is also interesting to note that Tps-system transported proteins have been associated with heme binding and utilisation functions (Zambolin et al. 2016).

The FhaC domain is identified in the differentially expressed *Trichodesmium* FHOM genes *Tery_3489*, *Tery_3487* (Fig. 5.22) and the non-differentially expressed *Tery_3266*. All three genes also contain the polypeptide-transport-associated (POTRA) domain. Gene *Tery_1533* (annotated as ‘polypeptide-transport-associated’ and not ‘filamentous hemagglutinin outer membrane protein’) also contains the POTRA and FhaC domains and is identified as differentially expressed in the transcriptome.

The 5 genes have mixed expression profiles across the 4 treatments indicating a response to different Fe cues either from dust, FeCl₃ or both (Fig. 5.22). Whether these genes might be involved in adhesion, Fe acquisition, or both, they seem to be significant and require further investigation.

The *Tery_08xx* gene cluster spans 8 genes (*Tery_0843-Tery_0850*) (Table 5D) all of which are differentially expressed in response to FeCl₃ and Dust (Fig. 5.23), with 5 being amongst the top 20 most differentially expressed genes identified in the presented analysis.

*Tery_0848* and *Tery_0849* are annotated as cell surface proteins while *Tery_0843* is recognised by the UniProtKB Automatic Annotation pipeline as a membrane protein. In addition, *Tery_0844* is predicted to contain an iron-sulphur binding site and *Tery_0845* belongs to the heme oxygenase superfamily. Therefore, some of these properties indicate a possible association with cell membrane systems and/or the Fe metabolism of *Trichodesmium* but there is no sufficient evidence to suggest an involvement in Fe acquisition. Further investigation is required to characterise the function of these proteins.
Surprisingly, the fifth gene in this cluster (*Tery_0847*) encodes for MetE (5-methyltetrahydropteroyltriglutamate/homocysteine S-methyltransferase) which catalyses the transfer of a methyl group to and form methionine. This vitamin B12-independent methionine synthase was previously identified to be inactivated under oxidative stress in *E. coli* (Hondorp & Matthews 2004). In addition, MetE functions as a replacement to MetH (B12-dependent methionine synthase) under vitamin B12-limitation but requires higher zinc and nitrogen concentrations in diatoms (Bertrand et al. 2013). *Trichodesmium* possesses two *metH* genes (*Tery_1073, Tery_2492*) which are downregulated in Fe+ compared to Fe- (opposite regulation to *metE*) but these are not significantly different between [Dust] and Dust+. It is difficult to suggest whether expression of the *metE* gene is regulated by the cell oxidation state, a dust-associated stimulus or a different cue associated with cellular contact to desert dust. Further understanding the controls on its expression, and the metal requirements for the protein’s production in *Trichodesmium*, is required to investigate its potential as a biomarker of dust-Fe availability in situ.

![Figure 5.23](image-url)

**Figure 5.23: Representative expression across the Tery_08xx cluster.** The expression of *Tery_0843* is higher in the Dust+ (dark brown) and Fe+ (yellow) treatments compared to [Dust] (light brown) and Fe- (green) respectively. Error bars represent standard deviations from the mean of 3 biological replicates.
Table 5D: The *Trichodesmium Tery_08xx* operon components. Features of interest indicate predicted conserved domains, subcellular localisation or metal binding sites (UniProtKB Automatic Annotation pipeline).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Annotation</th>
<th>Size (bp)</th>
<th>Features of interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tery_0843</td>
<td>hypothetical protein</td>
<td>609</td>
<td>5 alpha-helical transmembrane regions</td>
</tr>
<tr>
<td>Tery_0844</td>
<td>putative iron-sulfur cluster-binding protein</td>
<td>876</td>
<td>[2Fe-2S] cluster binding site</td>
</tr>
<tr>
<td>Tery_0845</td>
<td>TENA/THI-4 protein</td>
<td>705</td>
<td>Heme oxidase superfamily</td>
</tr>
<tr>
<td>Tery_0846</td>
<td>hypothetical protein</td>
<td>459</td>
<td>-</td>
</tr>
<tr>
<td>Tery_0847</td>
<td>5-methyltetrahydropteroylglutamate/homocysteine S-methyltransferase</td>
<td>2235</td>
<td>methionine synthesis</td>
</tr>
<tr>
<td>Tery_0848</td>
<td>cell surface protein</td>
<td>183</td>
<td>3 zinc-binding regions</td>
</tr>
<tr>
<td>Tery_0849</td>
<td>cell surface protein</td>
<td>573</td>
<td>-</td>
</tr>
<tr>
<td>Tery_0850</td>
<td>hypothetical protein</td>
<td>141</td>
<td>-</td>
</tr>
</tbody>
</table>

5.3.9 Regulatory DNA

*Trichodesmium* with 40% of its genome comprised of non-protein-coding DNA both in culture and in natural populations, is unique amongst cyanobacteria which as a group have an average of 15% (Larsson et al. 2011; Pfrendt et al. 2014; Walworth et al. 2015). Non-coding RNA that makes up about 80% of it (Walworth et al. 2015) could have an important regulatory role that provides *Trichodesmium* with an advantage in the fluctuating nutrient conditions of its natural environment.

Group II Introns function as ribozymes (catalytic RNA) and mobile genetic elements capable of self-splicing and inserting themselves in the DNA, resulting in interruption of their host genes. Most possess an intron encoded protein (IEP) that forms maturases, to aid splicing, and reverse transcriptases (RT). Open reading frame (ORF)-less group-II introns also occur and their splicing potentially requires the function of trans-encoded maturases (Meng et al. 2005).

*Trichodesmium* harbours 17 group-II introns which interrupt 11 genes, the highest number known to date for bacteria. Only 25% of bacterial genomes are known to contain at least 1 such element. In addition, despite their rarity amongst bacteria, 10 out of the 17 group II introns found in *Trichodesmium* are ORF-less (Pfrendt et al. 2014).
Here, we identify that 7 of the 11 group II intron host genes are differentially expressed across our experimental treatments. We suggest that the importance of these genes in the Fe response of *Trichodesmium* could be vital to provide necessary increased regulatory complexity in a fluctuating environment with ephemeral Fe supplies. The possibility that differential splicing of these introns occurs has to be addressed through further analysis.

**Twintron:** One of the elements mentioned above forms a twintron, an arrangement where one intron resides within another intron. This is yet another rare genetic phenomenon found in *Trichodesmium*’s rich regulatory toolkit. It interrupts gene *Tery_4732* which is annotated as RNase HI but the mature/spliced mRNA product appears to be a member of a conserved hypothetical protein domain seemingly linked to a function in RNA metabolism (Pfreundt et al. 2015). Annotated gene *Tery_4730* is located within the host intron and *Tery_4731* within the inner intron (Fig. 5.24).

In the RNAseq analysis performed here we observe that the twintron host is differentially expressed across the experimental treatments (Fig. 5.25). Expression of the first part of the host gene before the 5’ splice site of the host intron is increased in the Fe+ and Dust+ treatments compared to Fe- and [Dust] with a higher difference stimulated by additions of dust (2.3 times increase) compared to FeCl₃ (1.4 times increase). The same pattern is true for the host and inner intron. The second part of the host gene is expressed at much higher levels than the rest of the twintron region (5-13 times than the first part of the host gene) and expression is increased in the Dust+ and [Dust] cultures compared to Fe+ and Fe-.

Unequal expression levels are also observed between the two parts of the host intron (5’ side has 4-7 times higher reads).

Variability in splice site spanning reads is also observed. An increased number of reads covers the 5’ host gene-host intron splice sites in the Dust+ and Fe+ compared to the [Dust] and Fe- (only the former is statistically significant). Similarly, to the unequal transcription profiles between the upstream and downstream regions of the host gene the splice site spanning regions vary in coverage. The 3’ host gene-host intron splice site is spanned by a higher level of reads (1.3- 3 times) compared to the 5’ splice site and the expression varies between treatments so that Dust+ and [Dust] have increased coverage compared to Fe- and Fe+. There are no statistically significant differences in the read coverage of the host intron-inner intron splice sites but there is significantly reduced (2.7-7 times) coverage in 3’ compared to the 5’ site.
Figure 5.24: Representative expression profile along the *Trichodesmium* twintron arrangement. Read coverage from treatment Fe+ across the forward and reverse DNA strand is presented mapped against the annotated *Trichodesmium* genome (ARTEMIS genome browser) The inner intron (black), host intron (grey) and host gene (blue) are indicated along the expression profile.
Figure 5.25: Expression of the *Trichodesmium* twintron across the experimental treatments. The expression of the inner intron (black), the host intron (light grey) and host gene (dark grey) differs between the experimental treatments: Fe- (green), Fe+ (yellow), [Dust] (light brown) and Dust+ (dark brown), and along the length of the DNA arrangement. Error bars represent the standard deviations of three biological replicates.
Diversity generating retroelement: A highly abundant 265 nt diversity generating retroelement functioning by mutagenic homing was previously identified in the non-coding fraction of the *Trichodesmium* genome (Pfreundt et al. 2014). Its analysis revealed 12 genes targets (Pfreundt et al. 2014) of which all, with one exception, are not differentially expressed in our sequencing results. The genes are expressed at low levels and no consistent SNPs appear compared to the *Trichodesmium* genome. The differentially expressed gene *Tery_1727* has higher expression of its 5’ region in the Dust+ treatment, similar to an unannotated region preceding this gene.

The 265 bp retroelement itself (2632789r-2632524r) is also more highly expressed in Dust+ (Fig. 5.26). We suggest that expression of the retroelement might not be controlled by the same factors as expression of its target genes. A combination of stimuli could be responsible for co-expression and diversity generation through this mechanism.

![Figure 5.26: Expression of the *Trichodesmium* diversity generating retroelement.](image)

*Figure 5.26: Expression of the *Trichodesmium* diversity generating retroelement.* This 265 bp fragment is differentially regulated across the Fe- (green)-, Fe+ (yellow), [Dust] (light brown) and Dust+ (dark brown) treatments with high expression level indicated in the latter compared to the other conditions. Error bars represent standard deviations of the means of three biological replicates.
An overview: Although the analysis of non-coding/regulatory RNA was beyond the scope of this study the results indicated its central role in Trichodesmium’s Fe-stress/dust-encounter response. Group II intron host genes, the Trichodesmium twintrón region and the previously identified diversity generating retroelement (Pfreundt et al. 2014; Pfreundt et al. 2015) were all differentially expressed across the four growth conditions. A more extended analysis is required to indicate the specific nature of non-coding RNA controls on the transcription of genes involved in Trichodesmium’s Fe metabolism.

5.3.10 Other metals

As mentioned previously, in addition to gene expression differences stimulated by Fe, as both Dust and FeCl₃, certain genes demonstrate an expression profile where in both the Dust+ and [Dust] cultures expression levels are higher than in Fe+ and Fe-. This is likely stimulated by soluble (i.e. capable of passing through the dialysis tubing) chemicals from dust, potentially including trace metals such as manganese, cobalt and zinc. These metals could be presenting secondary nutrient limitation challenges to growing Trichodesmium cells in no-dust added treatments. To identify effects of manganese, zinc or cobalt to the Trichodesmium molecular response across the experimental treatments the expression of genes involved in their uptake is reviewed.

Manganese: Adjacent to the Fe transporter futBC (Tery_3222- Tery_3223) genes, Tery_3218 and Tery_3219 are homologous to the Anabaena all3575-alr3576 which are regulated by manganese (Huang & Wu 2004) and the manganese transporter MntAC (sll1599-sll1598) of Synechocystis (Bartsevich & Pakrasi 1996). The mntA homologue encoding for the ATP binding component of the transporter is differentially expressed with increased transcription in the non-dust (Fe-, Fe+) compared to dust (Dust+, [Dust]) treatments. This suggests that manganese might be released from dust and affect the expression of genes and processes differentially to that of the Fe- and Fe+ treatments.

Zinc: Homologues for the zinc transporter znuAB of Anabaena (all0833-all0832) are encoded by Tery_4951-Tery_4952, forming the putative periplasmic zinc binding protein, and ATPase respectively. The neighbouring Tery_4953 is related to an ABC-transporter permease. None of these genes show differential expression in the conditions tested here.

Cobalt: Genes Tery_2781-2782 encode for the cobalt transporting system, composed of the CbiQ inner membrane permease and the cobalt binding CbiM (homologues of Anabaena alr3945-alr3943). The ATP binding subunit CbiO is not annotated in the Trichodesmium
genome but downstream of the \textit{cbiQ/cbiM} genes, \textit{Tery\_2780} is homologous to the Anabaena CbiO gene, \textit{alr3949}. None of these genes are differentially expressed in the presented transcriptome. A fourth gene annotated as ‘cobalt transport protein’ \textit{Tery\_4901} is not identified as differentially expressed either.

5.4 Conclusions and further discussion

Despite \textit{Trichodesmium’s} role in the global biogeochemical cycles, little is known about its nutrient acquisition strategies. In particular, mechanisms employed by this organism in handling environmentally relevant nutrient sources, i.e. desert dust, have only recently started to be considered (Rubin et al. 2011). It has been demonstrated that environmentally collected \textit{Trichodesmium} puff colonies handle dust as a unit exhibiting traits of multicellularity, and that their contact with dust particles is potentially important for Fe acquisition. With the aim to better characterise \textit{Trichodesmium’s} ability to extract Fe from Saharan desert dust we used controlled laboratory conditions to grow \textit{Trichodesmium erythraeum} IMS101 in its presence, either when physically separated or by allowing direct contact. We compare the differences observed to those stimulated by additions of the conventional culture study Fe source FeCl$_3$-EDTA to a no added Fe background. Results indicate an advantage in utilising Fe to cells permitted to come in contact with dust. Therefore, Fe acquisition from particles likely exploits an as yet unspecified cell surface process.

To further investigate the Fe utilisation tools available to \textit{Trichodesmium} the transcriptomic response of cells grown in contact or separated from dust, with FeCl$_3$-EDTA (FeCl$_3$) or under Fe-deficient conditions was analysed. Dust-added cultures were found to be transcriptomically more similar to FeCl$_3$-added than Fe-stressed cultures which share expression features with the dust-separated cultures. Several patterns of gene expression were identified, including coordinated and unique changes by dust and FeCl$_3$-additions (compared to the dust-separated and Fe deficient conditions respectively).

Previously recognised Fe-limitation marker genes are either affected by both dust and FeCl$_3$ or only the latter. This might be the result of increased available Fe concentrations, the provision of uncomplexed rather than particulate Fe, or the absence of additional nutrients which are available in dust, in the Fe+ treatment. Regardless of which are the underlying factors, results raise the need to apply caution when using these biomarkers in the field.
where Fe concentrations are low and the major Fe source is Aeolian desert dust deposition. Based on the presented analysis, the petE-petJ genes of the plastocyanin- cytochrome c\textsubscript{553} switch are the most well suited for use as biomarkers of the cellular Fe state.

To summarise the results of the transcriptomic analysis, gene ontology (GO) annotations were used to investigate the involvement of differentially expressed genes in various processes. Processes like photosynthesis, N\textsubscript{2} fixation, oxidative phosphorylation and Fe transport were reviewed in detail.

The process of photosynthesis was significantly affected and changes were observed in response to both Fe sources- dust and FeCl\textsubscript{3}- simultaneously as well as by each addition separately. An effect on photosynthesis is to be expected as components including photosystems, electron transporters and photosynthetic pigments require Fe to be assembled and correctly function. In addition, when comparing the general response of genes for PSI/PSII subunits in Fe\textsuperscript{+} to Fe\textsuperscript{-} cultures, a general upregulation of PSII and downregulation of PSI genes is observed during Fe-deficiency. This conservation strategy previously identified to occur at the protein level minimises utilisation of the Fe expensive PSI (Snow et al. 2015) whose light harvesting efficiency is increased by upregulation of isiA (Bibby et al. 2001). Interestingly, cpcG2 which encodes for the CpcG phycobilisome paralogue that associates with PSI preferentially to PSII (as opposed to cpcG1) (Kondo et al. 2007; Deng et al. 2012) is also upregulated in Fe\textsuperscript{-} compares to Fe\textsuperscript{+}. We suggest that for the same reason as IsiA this functions to maximise efficiency of numerically decreased PSI complexes.

Although nitrogenase encoding genes, do not appear to be downregulated under Fe depleted conditions, as expected based on previous studies (Snow et al. 2015; Shi et al. 2007), expression of genes for enzymes nitrate/nitrite reductase and glutamate dehydrogenase involved in the organism’s nitrogen metabolism are affected. This could be the result of the N-limited state of cells under Fe-deficiency due to lower nitrogenase activity or the enzymes’ Fe content.

Furthermore, amongst the hypothesised Fe transport genes, the periplasmic binding proteins of the Fe\textsuperscript{3+} and siderophore uptake systems futA and fhuD, are FeCl\textsubscript{3} regulated. Although \textit{Trichodesmium} also possesses homologues to the other members of the Fut transporter (futB, futC) it does not appear to encode homologues of the Fhu permease (fhuB) and ATPase (fhuC) components. This could be the result of divergent protein
components to complement FhuD in *Trichodesmium* or reduction and dissociation of Fe from hydroxamate siderophores once in the cell periplasm with subsequent transport across the plasma membrane through a different uptake system (e.g. FeoAB or FutABC).

The reductive mechanisms involved are also uncharacterised but possibly involve the previously suggested superoxide-mediated reduction of Fe (Roe & Barbeau 2014; Hansel et al. 2016). In addition, *Synechocystis* has been suggested to utilise electrons exported by the terminal oxidase ARTO (Kranzler et al. 2014). Based on the presented differential regulation of this membrane complex at the transcriptomic level, we suggest that reduction of Fe by ARTO-exported electrons could also be employed by *Trichodesmium*.

Despite Fe being reduced prior to its uptake by several microorganisms, oxidation to Fe$^{3+}$ by multicopper oxidases (MCO) and uptake through specialised Fe$^{3+}$ transporters also occurs. We show the presence of an MCO-like gene in the *Trichodesmium* genome that is differentially expressed in response to FeCl$_3$ and could potentially be involved in *Trichodesmium*’s Fe uptake pathway.

Most of the information currently available about *Trichodesmium*’s Fe uptake system, as inferred by homology to proteins of other organisms, describe transport of Fe through the inner membrane. Interestingly, despite the presence of homologues to the TonB machinery (TonB, ExbBD) involved in siderophore uptake, *Trichodesmium* lacks known outer membrane TonB-dependent transporters (TBDT). It is possible that in this organism TBDTs are divergent to the known proteins of other cyanobacteria.

We identify an array of differentially expressed genes predicted to form large, calcium-binding proteins (annotated as Haemolysins-type calcium binding region domain proteins-HTCaB) likely to be localised at the cell outer membrane. These proteins have peptidase-like domains but could also be involved in adhesion and motility. Considering the involvement of superoxide in Fe uptake (Roe & Barbeau 2014), the interconnection of the ROS and Calcium cellular signalling systems (Gordeeva et al. 2003; Brookes et al. 2004; Yan et al. 2006; Singh & Mishra 2015) is also an interesting point to consider during further study of the HTCaB proteins. In addition, another set of possibly extracellular proteins that might be involved in adhesion and heme utilisation, annotated as filamentous hemagglutinin outer membrane domain proteins, has gene members displaying strong differential regulation across the experimental treatments. Lastly, the Tery_08xx cluster is a
novel set of genes with an apparent key role in *Trichodesmium*’s Fe-deficiency response and includes members predicted to be localised at the cell surface.

Despite the lack of information regarding *Trichodesmium* outer cell membranes, it is potentially a key cellular component playing a critical role in the organism’s life style. The *Trichodesmium* cells interact not only with nutrient substrates, but also with each other to form filaments and colonies in the environment, and with a diverse community of epibionts. Each of these features is significant to *Trichodesmium*’s success in the world oceans. Therefore, further study of the putative extracellular proteins identified is necessary.

The absence of an established system for genetic manipulation of *Trichodesmium* limits greatly the tools available for further study of proteins and genes. It would be beneficial for future research to attempt an optimisation of growth conditions and explore possible options for genetic engineering of the organism. Additionally, proteins can be heterologously expressed in other organisms (Chapters 3 and 4) to assist their functional characterisation.

In this study, changes in gene transcription have been used to better characterise the organism’s response to different Fe regimes. It is important to recognize that in addition to transcription several other controls mediate cellular activity. For example, pre-translation processes also include the interference of non-coding RNA. This appears to be particularly important for *Trichodesmium* whose non-coding genome proportion far exceeds the bacterial average (Larsson et al. 2011; Pfreundt et al. 2014). Indeed, we observe differential expression of the majority of group II intron hosts and the recently discovered twintron arrangement found in its genome (Pfreundt et al. 2014; Pfreundt et al. 2015). Finally, a diversity generating retroelement is stimulated by dust despite the fact that its target genes are expressed at low levels and are stable across the experimental treatments. We speculate, that its high non-coding genome content provides *Trichodesmium* with an additional regulatory capacity and flexibility, which is advantageous in a highly variable environment where nutrients fluctuate over time and space.

With a large fraction of *Trichodesmium* genes being miss- or unannotated it is likely that a wealth of information regarding its physiological adaptations to a nutrient deficient niche are disregarded. This information could provide a window to understanding the evolution of adaptive strategies to a broader range of marine primary producers and assist in the
development of tools for detection and prediction of nutrient limitation in the ocean. Therefore, future efforts will be directed to a better characterisation of the *Trichodesmium* genome primarily though bioinformatic but also using analytic targeted techniques of gene/protein function in order to accommodate a more thorough understanding of the increasing pool of omic-technique derived data.
Chapter 6
Synthesis

6.1 Introduction

The research presented in the previous chapters aimed to better characterise *Trichodesmium* iron (Fe) and phosphorus (P) utilisation mechanisms and to provide a better understanding of the nutrient stress responses of this organism using a series of laboratory studies, involving both targeted and holistic molecular approaches.

In Chapter 3, the ability of *Trichodesmium* to grow on the reduced inorganic source of P, phosphite (PO$_3^{3-}$) was assessed. A gene cluster with homology to the PO$_3^{3-}$ utilisation operon, *ptxABCD*, of pathogenic bacteria and *Prochlorococcus marinus* str. MIT 9301 was examined. Heterologous expression in *Synechocystis* sp. PCC 6803 (*Synechocystis* hereafter) was used to characterise its function, and its regulation by different P regimes was studied by quantitative RT-PCR (qPCR).

The technique of heterologous expression was also used in Chapter 4 to address the ambiguous function of protein Tery_3377, also referred to as IdiA. Although the protein is assumed to be involved in Fe transport it is equally similar to the *Synechocystis* FutA1 and FutA2 paralogues which are differentially localised and perform dissimilar functions.

Finally, acquisition of Fe from Saharan desert dust by *Trichodesmium* and the role of cell surface contact for this process was assessed through a physiological experiment. Transcriptomic changes stimulated by Fe-limitation or provision of dust were analysed in an attempt to explore the organism’s adaptive strategies to environmentally relevant Fe fluctuations.
6.2 Main findings and concluding remarks

6.2.1 Better characterisation of nutrient uptake systems

*Trichodesmium*’s niche spans vast nutrient limited areas of the low latitude ocean (Sohm et al. 2011), therefore the organism is hypothesised to possess efficient Fe and P utilisation systems that target several forms of these nutrients (Fig. 6.1). As *Trichodesmium* is currently not genetically tractable, most of what is presently known about the molecular control of such pathways is inferred from homology to characterised proteins of other organisms.

Work performed in the context of Chapters 3 and 4 demonstrated the efficacy of heterologous gene expression as a technique for characterising the function of *Trichodesmium* proteins. The action of two nutrient uptake/utilisation pathways of this organism was elucidated: The Fut system for uptake of Fe$^{3+}$ across the plasma membrane and the Ptx system for uptake and utilisation of PO$_3^{3-}$ (Fig. 6.1). These constitute the first direct evidence for the functioning of nutrient uptake pathways in this organism.

**Phosphite, an unlikely source of P in the ocean**

As described in Chapter 3 the *Trichodesmium ptxABCD* cluster is responsible for uptake and utilisation of PO$_3^{3-}$ for growth. Essential for this pathway are both the ABC transporter (*ptxABC*) and the enzyme phosphite dehydrogenase (*ptxD*). These genes are also identifiable in published metagenomic and metatranscriptomic datasets from the North Atlantic and West Pacific (Fig. 6.2). Their prevalence and expression *in situ* indicates that their role stretches beyond that displayed in the controlled environment of laboratory cultures.

The new data presented adds to recent evidence which suggest that oceanic P exists in more than just its fully oxidised $+$5 valence state in the marine environment, and that valence state $+$3 P is available to marine cyanobacteria both as organic phosphonates (Pn) and inorganic PO$_3^{3-}$ (Martínez et al. 2012; Dyhrman et al. 2006; Van Mooy et al. 2015). One source of Pn to the ocean appears to be microorganisms themselves (Dyhrman et al. 2009; Pasek et al. 2014; Van Mooy et al. 2015). This could also be true for PO$_3^{3-}$ as Van Mooy et al. (2015) showed its production by water column sample incubations and *Trichodesmium* colonies as well release of the majority of P reduced by the latter.
Chapter 6: Synthesis

Figure 6.1: Iron and Phosphorus uptake by *Trichodesmium*. Subunits with homology to known Fe (top) and P (bottom) uptake systems are indicated. The P uptake pathways include the Phn organophosphonate (Pn) transporter and C-P lyase (orange), the APase for catalysis of organophosphates (green), the high affinity phosphate (PO₄³⁻) transporter Pst (yellow) and the Ptx transporter shown to be responsible for PO₄³⁻ utilisation in Chapter 3 (grey). Subunits involved in utilisation of Fe include organic ligand-Fe transporters of the TonB/ExbBD system and periplasmic component FhuD (orange). We hypothesise a role of ARTO in reduction of organic ligand bound Fe (orange). Fe³⁺ might be transported directly through the Fut system which includes the Fe-binding, periplasmic protein, FutA (Chapter 4) (yellow) or reduced by ARTO/FRE (Chapter 5) (yellow). Similarly, Fe²⁺ might be taken up by the Feo system (green) or oxidised by a MCO protein before uptake (Chapter 5) (green). Possible outer membrane HTCaB and FHOM.
Although the purpose of biological PO$_3^{3-}$ production is unknown, it could simply be an intermediate product of cellular Pn production and degradation pathways (Pasek et al. 2014). Based on the compound’s antifungal qualities (Griffith et al. 1989; Karl 2014), possible use of PO$_3^{3-}$ as an antibiotic defence by its producers is also suggested. In the case of *Trichodesmium* this could extend to protection of the entire colony consortia and/or keeping a balanced epibiont community. Additionally, if the genetic toolkit for accessing this source of P is not universal in the ocean, the specificity of nutrient exchange between *Trichodesmium* and specialist PO$_3^{3-}$ utilising epibiont species could benefit the symbiotic relationship.

Regardless of the reasons behind biological PO$_3^{3-}$ production, its presence in the ocean renders the existence of specialised uptake pathways (PtxABC transporter) beneficial for marine organisms. Intracellularly, although C-P lyases or APases, used for catalysis of Pn and organophosphate respectively, are also capable of driving the oxidation of PO$_3^{3-}$ to PO$_4^{3-}$ (Metcalf & Wanner 1991; Yang & Metcalf 2004) (Fig. 6.1), NADPH/NADH formation (Costas et al. 2001) during PtxD activity makes this reaction energetically favourable. Therefore, the latter is more likely to be the preferred pathway for transforming PO$_3^{3-}$ to PO$_4^{3-}$ which will be consequently used for the cells’ metabolic reactions.

**A better characterisation of the Fe regulated IdiA**

The ability of *Trichodesmium* to take up Fe$^{3+}$ has not been clearly established in previous studies. Although reduction prior to Fe uptake is a major step of Fe acquisition by this organism, evidence indicates that free Fe$^{3+}$ is also directly available (Roe & Barbeau 2014). However, the pathways involved have not been described at the molecular level.

Protein Tery_3377 is considered a periplasmic Fe$^{3+}$ binding protein, part of an ABC uptake transporter, but is confusingly annotated after the *Synechococcus* sp. PCC 6301/PCC 7942, intracellular IdiA, involved in protection of PSII from oxidative stress (Michel et al. 1996; Michel et al. 1998; Exss-Sonne et al. 2000), due to their homology. They are both similar to the pathogenic bacteria Fe-binding FbpA part of the Fbp ABC-transporter for transport of Fe$^{3+}$ (Krewulak & Vogel 2008). *Synechocystis* has two paralogues of this protein termed FutA1 and FutA2 which appear to be differentially localised, performing dissimilar functions one potentially similar to the pathogenic bacteria transporter FbpA (FutA2) and the other to the *Synechococcus* PSII-associated IdiA (FutA1).
Figure 6.2: The prevalence /expression of the ptx gene cluster in marine cyanobacteria *Trichodesmium* and *Prochlorococcus*. Identified members of the cluster in published metagenomes and metatranscriptomes of *Trichodesmium* (summarised in Polyviou et al. 2015) (red) and *Prochlorococcus* (Martínez et al. 2012) (orange) mapped over the global distribution of \( \text{PO}_4^{3-} \) in the ocean (Karl 2014).

In Chapter 4, data to support the role of *Trichodesmium* Tery\_3377 as an \( \text{Fe}^{3+} \) transporter is presented (Fig. 6.1). The protein possesses a signal sequence that directs it to the periplasmic space through the twin arginine transporter (TAT) system. The purified protein binds \( \text{Fe}^{3+} \) with high affinity and the protein structure bound to \( \text{Fe}^{3+} \) is retrieved (Polyviou et al, manuscript in preparation). Finally, heterologous expression in *Synechocystis* futA2 mutants appears to relieve the compromised physiology (generated when \( \text{Fe}^{2+} \) is bound to the strong ligand ferrozine). We therefore propose the reannotation of Tery\_3377 as FutA2 (referred to it as Tery\_3377 for the purpose of this discussion to avoid confusion).

Despite the identified function of *Trichodesmium* Tery\_3377 as an \( \text{Fe}^{3+} \) transporter, the collected results raise questions about the exact function of the two *Synechocystis* paralogues. Disproportional growth reduction of the double mutant \( \Delta \text{futA1}/\Delta \text{futA2} \) compared to the pooled decrease of \( \Delta \text{futA1} \) and \( \Delta \text{futA2} \) growth, could be the result of accumulated pressures above the level at which they can be internally managed by the
organism, when two rather than one Fe-stress response proteins are deleted. However, this disproportional growth reduction could also be an indication of a complementary FutA1 and FutA2 function, in which case at least one of the two will be required for efficient growth under Fe-stress. Traces of GFP in the periphery of cells when fused to the FutA1 signal sequence is possible evidence towards a more intricate function of FutA1, with a hypothesised additional role in Fe uptake. This theory is also supported by the growth/photosynthetic physiology improvement of the Tery_3377 expressing ΔfutA1+ compared to ΔfutA1. Other justifications for this recovered physiology of ΔfutA1+ include a putative secondary function of Tery_3377 similar to that of the intracellular FutA1, or ‘boosting’ of Fe uptake (resulting from regulation of its expression by a constitutive promoter) favouring the ΔfutA1 independently.

These questions remain to be elucidated in more in-depth studies focusing on the functional characterisation of the Synechocystis paralogues. An interesting approach could be to in vivo track FutA1/FutA2 localisation, during the diel cycle and Synechocystis growth phase stages, both under Fe-replete and deplete conditions.

**Cell Surface- The importance for utilisation of dust-Fe**

The involvement of cell surface/outer membrane processes in Fe uptake is not currently well characterised in marine cyanobacteria. *Trichodesmium* is capable of utilising siderophores and dust for which passive diffusion through the cell membranes is unlikely (Achilles et al. 2003; Langlois et al. 2012; Rubin et al. 2011, Chapter 5). The dissolution and acquisition of Fe from dust and ferrihydrite appears to occur extracellularly and it is suggested that processes at the cell surface involving substrate adsorption on the cells is a necessary prerequisite for this process (Rubin et al. 2011). The latter was indicated through inhibition of Fe dissolution by entrapment of ferrihydrite within glass fiber filters. There are pitfalls to this technique as it is not possible to demonstrate whether dissociation of Fe is prevented because of the way that ferrihydrite is attached to the filters. Also the media used does not include any ligands such as EDTA to prevent re-precipitation of Fe immediately after its potential dissolution.

Knowledge of pathways for extracellular, cell surface-associated, Fe reduction/dissolution from ligands in marine cyanobacteria is limited. *Synechocystis* pili appear to facilitate electron export for Fe reduction (Lamb et al. 2014) while in *Trichodesmium* cell derived superoxide might be involved in the process (Roe & Barbeau 2014; Hansel et al. 2016). In addition, other than TonB dependent transporters (TBDTs) (Hopkinson & Morel 2009), no
known proteins facilitating cell surface adsorption/uptake of Fe/Fe ligands are characterised in marine cyanobacteria. In *Trichodesmium*, although periplasmic and inner membrane components the TBDT pathway (TonB and ExbBD) have been identified (Fig. 6.2), no obvious homologues of the outer membrane components are encoded in the genome.

In Chapter 5 the question of the role of cell surface processes to extracting Fe from dust was addressed in a simple and direct experimental set-up. Dust was either released within the culture flasks in contact with cells (Dust+), or separated by means of a dialysis membrane ([Dust]). The growth and photosynthetic physiology was compared to no Fe (Fe-) and FeCl$_3$-EDTA added (Fe+) cultures. The results revealed that when direct contact with dust is prevented, Fe limits *Trichodesmium* growth and photosynthetic efficiency.

The treatments were analysed using RNA sequencing (RNAseq) to look deeper into the molecular pathways underlying the growth and photosynthetic physiology response of the cultures to the different Fe regimes. Uncharacterised genes encoding for proteins possibly associated with the cell/outer membranes were identified based on conserved protein domains. Protein groups annotated after the ‘Haemolysins-type calcium binding region domain’ (HTCaB) and ‘Filamentous hemagglutinin outer membrane domain’ (FHOM) as well as an eight gene cluster (Tery_08xx) included some of the most significantly differentially expressed genes between the experimental treatments.

Proteins of the HTCaB group have recognised similarity to COG2931 (Ca$^{2+}$-binding protein) and Peptidase_M10_C (Peptidase M10 serralysin C terminal) conserved domains. These domains are encountered in extracellular metallopeptidases involved in proteo/hemo-lytic activity of pathogenic bacteria (Miyata et al. 1970; Boehm et al. 1990; Duong et al. 1992; Rose et al. 1995; Kuhnert et al. 2000; Christie-Oleza & Armengaud 2010; Massaoud et al. 2011). Roles in adhesion and motility have also been suggested (Brahamsha & Haselkorn 1996; Hoiczyk & Baumeister 1997; Pitta et al. 1997; Sánchez-Magraner et al. 2007). To our knowledge the role of these domains has not been previously discussed in cyanobacteria. The properties of adhesion, motility and lysis of extracellular substrates could be beneficial qualities for particulate Fe utilisation by *Trichodesmium*.

The same is true for the FHOM protein group which is also associated with virulence of pathogenic bacteria trough adhesive properties (Melvin et al. 2015). Such proteins are extracellular, exported from the cell through the two-partner secretion (TPS) system.
(Hodak et al. 2006). It is worth mentioning that TPS translocated proteins can also be non FHOM containing such as the *Haemophilus influenzae* HxA. Interestingly, this protein has a role in heamopexin utilisation by extracting its heme group which is then available for transport through the TonB-dependent transporter (TBDT) protein HxuC. It can be hypothesised that the FHOM proteins in *Trichodesmium* have evolved to perform similar functions.

It is interesting to speculate whether in association with TonB a diversified outer membrane-linked protease has evolved to perform both functions of extraction and translocation of Fe from ligands. Possible involvement of HTCaB and FHOM proteins remains to be examined.

Another group of uncharacterised genes with strong differential expression in the transcriptomic analysis, contains 8 genes (*Tery_0843- Tery_0850*) sequentially localised in the *Trichodesmium* genome. Two are annotated as cell surface proteins and one is predicted to be membrane associated but further analysis is required to identify possible association with extracellular processes.

Based on the transcriptomic results (as presented in Chapter 5), it appears that processes occurring at the *Trichodesmium* cell surface are fundamental in coordinating its response to desert dust additions, and further investigation is required to characterise the function of identified genes with potential relevance to this. Heterologous expression, as demonstrated in Chapters 3 and 4, could be used as an approach for their localisation and functional characterisation, while the metal binding capacity of proteins can be tested *in vitro*.

**The skeleton-frame of *Trichodesmium’s* Fe uptake system**

Through research performed in the context of this thesis, progress is made in elucidating *Trichodesmium’s* Fe uptake system. We suggest a working model (Fig. 6.1) through which putative outer membrane HTCaB and FHOM proteins act through lytic/adhesive properties to extract Fe from dust. Additionally, either in combination or independently to these processes, Fe dissolution via reduction by cell derived ROS (Roe & Barbeau 2014; Hansel et al. 2016) might be involved. However, no information regarding the molecular pathways behind this process is currently available.

Released Fe$^{2+}$ and Fe$^{3+}$, uncomplexed or bound to organic ligands, is subsequently transported across the outer membrane through unidentified TBDTs. In the periplasm, FhuD binds ligand-associated Fe and either leads it to an as yet unidentified inner
membrane permease (related to FhuB) or an electron source such as ARTO for its reduction to Fe$^{2+}$ (Fig. 6.1). Likewise, FutA2 binds Fe$^{3+}$ and either facilitates its transport across the inner membrane in combination with FutBC or releases it after its reduction (Fig. 6.1). Reduced Fe released in the periplasm will be subsequently transported by the FeoAB system.

Putative ferric reductase (FRE) and multicopper oxidase (MCO) subunits potentially confer redox activity contributing to Fe transformations either intracellularly, or at the exterior of the cell prior to its uptake, complementing processes of the hypothesised model.

### 6.2.2 Dust/FeCl$_3$ effects on photosynthesis

The most prominent changes identified to be coordinated by FeCl$_3$ and dust are related to photosynthesis. This could, to a large extent, be due to the better characterisation of photosynthesis compared to other cellular processes. Changes in response to one or more of the experimental treatments are identified for all the major photosynthetic electron transport chain components (Fig. 6.3).

Similar responses are induced by FeCl$_3$ and desert dust in genes of: the cytochrome subunits of PSII, cytochrome $c_{550}$ and $b_{559}$, the PSI subunits PsaA and PsaB binding the $F_X$ Fe-S centre, several phycobilisome proteins, the electron transporters plastocyanin and cytochrome $c_{553}$, one flavodoxin protein, the NADPH dehydrogenase NdhI binding the Fe–S clusters N6a and N6b, and components of the F$_1$F$_o$ ATP synthase complex.

Genes for several subunits of the phycobilisome antenna are commonly upregulated in the presence of dust and FeCl$_3$, as expected from the high Fe requirements of phycobilisome synthesis. However, interestingly, the opposite behaviour (upregulation in Fe-/Dust) is displayed by the gene encoding CpcG2 for which there is evidence indicating an involvement in organisation of a PSI phycobilisome antenna (Deng et al. 2012; Watanabe et al. 2014). Although the attachment of the phycobilisome antennae to PSII is generally well described, a justification for the nature of phycobilisome-PSI interaction was only recently published (Watanabe et al. 2014). It appears to involve a supercomplex of a double-dimeric (tetrameric) PSI complex which is associated with one to three rods of two hexameric phycocyanin units. As CpcG2 is mostly found in daytime fixing diazotrophs compared to species like *Crocosphaera* that fix N$_2$ in the dark, and the concentrations are elevated in heterocysts rather than vegetative cells, the authors suggest an important function related to N$_2$ fixation. The tetrameric PSI-phycobilisome supercomplex could be enhancing cyclic
electron transport (CET) as a way of reducing oxygen (detrimental to nitrogenase) evolution by PSII (Fig. 6.3). Our results can be explained by an N-depleted state of Fe- and dust-separated cultures stimulating upregulation of the supercomplex to allow more efficient N\textsubscript{2} fixation. It can be also suggested that formation of the supercomplex is an Fe-stress response strategy facilitating an increased light harvesting efficiency by the fewer PSI subunits (reduced due to their high Fe requirements relative to PSII) in a similar way as IsiA antennae proteins.

Figure 6.3: The hypothesised response of the photosynthetic electron transport (PET) chain to Fe limitation in *Trichodesmium*. Based on the direction of the expression change of the majority of genes for each subunit as identified in Chapter 5 of this thesis (Fe- vs Fe+ comparison). Photosystem I (PSI), PSII-phycobilisome, cytochrome C\textsubscript{553} (Cyt C\textsubscript{553}) and F\textsubscript{o}F\textsubscript{1} ATPase (ATPase) are downregulated under Fe-stress (red). Conversely, NAD(P)H dehydrogenase, photosystem II (PSII), cytochrome b\textsubscript{6}f (Cyt b\textsubscript{6}f), plastocyanin (PC) flavodoxin (Flvd) and ferredoxin (Frd) appear to be upregulated (green) under the same conditions. The PSI phycobilisome is hypothesised to be upregulated (green) based on increased expression of the *cpcG2* gene important for its organisation. Cyclic electron flow (CEF) around PSI is enhanced under low Fe (bold arrows).

Despite the similarities mentioned there are also differences between gene expression responses of *Trichodesmium* to FeCl\textsubscript{3} compared to dust. Genes that are only differentially expressed in response to contact with dust encode for the PSII transmembrane CP43 and CP47, additional subunits of the phycobilisome antenna, F\textsubscript{i}F\textsubscript{o} ATP-synthase, NADPH dehydrogenase and the ferredoxin NADP+ reductase. Genes only regulated by FeCl\textsubscript{3} include the D1/D2 of PSII, the membrane intrinsic subunits of cytochrome b\textsubscript{6}f, an additional flavodoxin, two ferredoxin proteins, and an additional component of NADPH
dehydrogenase. These differences might either be associated with the nature/quantity of Fe available in the two treatments or the presence of other chemical stimuli from dust. A general upregulation of the photosynthetic components in both the dust-separated and dust-added cultures compared to Fe- and Fe+ is also observed. It is difficult to indicate the stimuli responsible for these differences based on the provided data alone. Elevated concentrations of specific nutrients from dust (e.g. Manganese -See Sec. 5.3.10) could be responsible if limiting in the non-dust cultures. As discussed below, alterations of the general ratio of resources in the dust containing media could also be important.

6.2.3 Is it more than just the concentration of nutrients? The problem with biomarkers of iron limitation

Phosphorus and Fe colimitation has been identified both in field and laboratory studies (Mills et al. 2004; Walworth et al. 2016; Garcia et al. 2015). Garcia et al. 2015 has shown that in P-deficient conditions Trichodesmium and Crocosphaera that are Fe-challenged grow and fix N2 faster than when provided with an abundance of Fe and attributed this to a decrease of cell size upon colimitation. The hypothesis (‘Nutrient-Balance’ hypothesis hereafter) is that marine life has adapted to colimitations and combination additions of nutrients and will thus have more effective strategies to deal with such scenarios. This could result in inaccurate results from several field and laboratory studies attempting to elucidate the response of organisms to single nutrient deficiencies/additions.

This ‘Nutrient-Balance’ hypothesis is critical to the arguments presented in Chapter 5 regarding the in situ use of nutrient limitation biomarkers identified in laboratory stress experiments. Culture experiments almost exclusively utilise nutrient forms that are of little relevance to the natural sources introduced to the ocean system, and the resources are provided in excess. This sets the interpretation of results more manageable but less representative of the natural world. Therefore, the differential expression of molecular Fe-limitation biomarkers identified in culture experiments could be indicative of the nutrient’s limitation/abundance but also:

- The nutrient-balance
- Regulation by a specific form of nutrient provided in the experiment
- The presence of a large excess of that nutrient

The former two can set the biomarkers environmentally irrelevant. For example, although the usual supplement in Trichodesmium Fe+ cultures in laboratory Fe-stress experiments is Fe’ from FeCl3-EDTA, desert dust is the primary source of Fe to the open ocean and will also
deliver other nutrients such as P, zinc, cobalt, nickel, manganese, and copper (Duce et al. 1991; Okin et al. 2004). In addition, the provision of a large excess of the nutrient in cultures can result in identification of biomarkers that will lack the sensitivity often required for detection of milder nutrient fluctuations often encountered in situ. In Chapter 5 *Trichodesmium* cultures were supplied with either FeCl₃·EDTA or Saharan desert dust as a source of Fe. Several common gene expression responses were recognised for the two sources but differences were also obvious.

Furthermore, the expression levels of a gene are often highly growth-phase dependent. This can be monitored through the expression profile of a gene/protein in parallel to the growth curve of laboratory cultures and also the expression during the development of phytoplankton bloom events (Shi et al. 2007; Spungin et al. 2016). It is also worth remembering that a biomarker chosen and tested based on the physiology of one organism will not necessarily be representative of its behaviour in other members of the community. For example, although *Trichodesmium*’s abundance/distribution was repeatedly shown to follow dust deposition (Langlois et al. 2012; Langlois et al. 2008; Lenes et al. 2008; Moore et al. 2009; Fernández et al. 2010) the response of *Synechococcus* and *Prochlorococcus* to dust can be variable with negative/toxic effects indicated in some occasions (Langlois et al. 2012; Paytan et al. 2009).

The arguments stressed above can, depending on the experimental approaches, either present limitations to the use of molecular biomarkers as detectors of nutrient stress, or provide a platform for novel applications of these tools. It is theoretically possible, using biomarkers, to detect bioavailability of different nutrient sources- the nutrient status- and the progression stage of a bloom event in-time. The Nutrient-Balance disturbance of natural/anthropogenic events could also be followed after suitable tool optimisation.

We suggest the following pipeline to be followed from the identification to the utilisation of marine gene/protein expression biomarkers of nutrient limitation: (1) Transcriptomic/proteomic analysis with a gradient of environmentally relevant nutrient forms to identify potential biomarker targets. (2) Characterisation of the targets’ differential expression development over time through quantitative RT-PCR (qPCR). (3) In situ testing. Sampling and expression-detection sensors (currently under development) have the capability of monitoring expression of the identified biomarkers over the course of bloom events. (4) Use of biomarkers to detect nutrient-stress in the field. Finally, multiplicity of biomarkers
targeting concurrent identification of various nutrient stressors can provide a more comprehensive picture (Saito et al. 2014).

In Chapter 5 we identified that some commonly used biomarkers of Fe-limitation were not differentially regulated by dust at the time of sampling while expression differences were strong in response to FeCl₃-additions. These include genes for the PSI antenna IsiA, the flavodoxin protein IsiB and the Fe transporter Tery_3377 (FutA2). Commonly differentially expressed genes between the two Fe sources include the interchangeable electron transporters plastocyanin and cytochrome c₅₅₃ (3rd and 1st most differentially expressed genes respectively). The copper containing plastocyanin is known to replace the Fe containing cytochrome c₅₅₃ under conditions of Fe-stress in a behaviour referred to as the plastocyanin-cytochrome (P-C) switch (Wood 1978; De la Cerda et al. 2007). We suggest further testing of the P-C switch, for its application as a biomarker of Fe-limitation. Due to the copper dependent nature of plastocyanin, the simultaneous use of an associated copper-status biomarker and/or seawater copper concentration measurements may be necessary.

6.2.4 Non-coding RNA- A role in the response of Trichodesmium to dust

Data presented in Chapter 5 indicate a significant involvement of non-protein coding DNA in the Fe-limitation/Dust-addition response of Trichodesmium. The unexpectedly high proportion of non-coding genome in Trichodesmium (40%), far exceeding the average for other cyanobacteria (15%), has been recently discussed after the characterisation of its primary transcriptome (Pfreundt et al. 2014) which indicated the highest number of non-coding DNA transcription start sites (TSS) compared of all bacterial species currently analysed. Unusual non-coding RNA properties, including a bacterial record breaking number of 17 group II introns targeting 10 host genes, a group II twintron arrangement (the only known bacterial twintron to interrupt a protein coding gene), as well as a diversity generating retroelement (DGR) targeting 11 different genes, have been identified (Pfreundt et al. 2014; Pfreundt et al. 2015). These, properties are added to the previously identified high number of transposable elements in this organism (Larsson et al. 2011; Lin et al. 2011; Mikheeva et al. 2013). The non-protein coding fraction of the genome seems to also be maintained in the environment (Walworth et al. 2015). This finding leads to the conclusion that regulatory RNA is linked to the organisms’ lifestyle, possibly its cohabitation with other microorganisms and the nutrient fluctuating conditions it encounters.
In the transcriptome presented, 7 out of the 11 group II intron host gene regions are identified to be differentially expressed with the twintronic region displaying a very high positive response when dust is in contact with *Trichodesmium* cells. In addition, although no major differences in expression of the 12 DGR gene targets between treatments are obvious, the retroelement itself is again upregulated in response to contact of cells with dust. We suggest that *Trichodesmium*’s unusual genome features may be linked to episodic desert dust influx events in the environment, and potentially its success in the ocean.

### 6.3 Final discussions

The study of biology at the systems level through omic approaches is expanding rapidly and provides a wealth of new information. As demonstrated, it can be useful in identifying key genes, both studied and uncharacterised, involved in organismal metabolic processes. In addition, the tools currently available enable the decryption of environmental metagenomes and metatranscriptoms *en masse* in projects like the Global Ocean Sampling (GOS) and TARA Oceans project while sequencing of single cells is opening a whole new suit of possibilities for marine research.

However, data generation currently far exceeds our analytic capabilities and is, in my opinion, the limiting step to a better characterisation of biological systems and processes. Taking *Trichodesmium* as an example, the function of the majority of the genome is unresolved with several unannotated, and a high incidence of miss-annotated, proteins often setting the generated results incomprehensible. Future efforts will need to be directed towards better bioinformatic and targeted gene characterisation of this species.

This characterisation, as demonstrated, can utilise state of the art genetic modification techniques even in a genetically intractable species such as *Trichodesmium*. This organism has an unusually large ‘deck of genes’ identification of which is not only important to understanding its possible response options in the future of ocean changes, but also to discover biotechnologically interesting potentials. As an example, the *Synechocystis* PtxABCD+ mutant able to grow on reduced inorganic P (Chapter 3, Polyviou et al. 2015), is currently under optimisation for use in bioreactors. Specialised growth on PO$_3^{3-}$ can promote culturing specificity and inhibit growth of contaminating species.

In the meanwhile, the presence and expression of this gene cluster in the environment is strong evidence for significant PO$_3^{3-}$ availability in the ocean, something that has not yet been demonstrated, with environmental concentrations and quantified fluxes of reduced P
(both organic and inorganic) currently lacking. The detection limits of analytical chemistry techniques for P but also for Fe may lead to misguided interpretations of nutrient availability to primary producers. The biological approach for identifying the presence of available nutrient pools through the use of biomarkers comes with its own limitations. Coupling of chemical measurements with biomarker detection through the development of integrated biochemical sensors is an approach with high potential, able to combine the advantages of both methods and deliver a more representative picture of the presence/availability of nutrients as well as their interaction with phytoplankton. Monitoring changes in the nutrient/production status of the world oceans, through autonomous *in situ* marine sensors, can indicate disturbances in the system over space and time and facilitate the prediction of biogeochemical cycle functioning and associated climatic changes.
Table A1: Primer sequences used for genetic modification.

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Purpose</th>
<th>Method</th>
<th>Construct</th>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Heterologous expression of Trichodesmium ptx genes in Synechocystis</td>
<td>pFLAG</td>
<td>ptxABCD</td>
<td>F</td>
<td>ATGCCATATGATCTCTAAATACGCGGTTGAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R</td>
<td>ATGCCGATCTCTAATAGCTGTTAAGCGTCTTTGA</td>
</tr>
<tr>
<td>4</td>
<td>Generation of Synechocystis knockout mutants</td>
<td>OLE-PCR</td>
<td>futA1</td>
<td>P1</td>
<td>CCAACTTCCTAGTCCCTGAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P2</td>
<td>ACATTATGCTGGTGCCTCACTGCACGGGATAAAGCTTGGACC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P3</td>
<td>AAACCTCAATCCTGATGCATAGGTGCAATGGCCATGTTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P4</td>
<td>GCCATTACGAAGGGGTTGGACG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P5</td>
<td>CGGACATTTTCAGCAGCTGTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P6</td>
<td>GGGCTTGCTATTGCCTGGCTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>zeo_F</td>
<td></td>
<td>GCAATGAGCGCAGCAACCTGAATATTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>zeo_R</td>
<td></td>
<td>ATGGCTGCAAGGCCATTAGTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OLE-PCR</td>
<td>futA2</td>
<td>P1</td>
<td>AAGGCGATGTAGTAAAGGCAAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P2</td>
<td>AAAGTTGCTGGCGAGCTGTTCCTTCCGTTATACCACTAAGGACAGTGAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P3</td>
<td>CGATGCTGAGCGGAGGGGGCCGTAAACCGATTCGCTGTTATAATG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P4</td>
<td>GGGATTCTTTGGCAGGAGT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P5</td>
<td>AGCGATACCAAATGGAAGGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P6</td>
<td>CGTGTCATTTACCCTCAGGTT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>com_F</td>
<td></td>
<td>TACGCCGAGCCCTGGGCGGACCTTT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>com_R</td>
<td></td>
<td>TTACGCCGCCCTGGGCGACCTAG</td>
</tr>
<tr>
<td></td>
<td>GFP-protein signal sequence co-expression in Synechocystis</td>
<td>OLE-PCR</td>
<td>futA1ss-GFP</td>
<td>F</td>
<td>CAAATACATATAAGGAAATATTAAACCAAATGTCACAAATGGTATCGCTT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R</td>
<td>TGAACGCTTCCTCGGCTTACGATCCCTGTTGCTGTCGCGAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>futA2ss-GFP</td>
<td>F</td>
<td>CAAATACATATAAGGAAATATTAAACCAAATGACAAATAGATTTCGGGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R</td>
<td>TGAACGCTTCCTCGGCTTACGATCCCTGTTGCTGTCGCGAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>337Tss-GFP</td>
<td>F</td>
<td>CAAATACATATAAGGAAATATTAAACCAAATGACAAATAGATTTCGGGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R</td>
<td>TGAACGCTTCCTCGGCTTACGATCCCTGTTGCTGTCGCGAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>torAss-GFP</td>
<td>F</td>
<td>CAAATACATATAAGGAAATATTAAACCAAATGACAAATAGATTTCGGGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R</td>
<td>TGAACGCTTCCTCGGCTTACGATCCCTGTTGCTGTCGCGAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>psbA2ss-GFP</td>
<td>F</td>
<td>GGAATATCGAATGACAAATAGATTTCGGGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R</td>
<td>TGAACGCTTCCTCGGCTTACGATCCCTGTTGCTGTCGCGAT</td>
</tr>
<tr>
<td>Gene</td>
<td>Source</td>
<td>Sequence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>psbA2ds</td>
<td>R</td>
<td>TTGGTTATAAATCTTTAACGTGATTTGTCGATG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>CGATGAGTGCCAGGGCCGGGGGGCTGAAAATTTCTTTGTTGTATGACACCTG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>CTGGTGAAAGCCTGGCGG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kan</td>
<td>F</td>
<td>TACCCGGAAGCCCTGGGCCCAACTTT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>TACCCGGAAGCCCTGGGCCCAACTTT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GFPss</td>
<td>F</td>
<td>ATGCGTAAAGGCGAAGAGCT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>AAGATTGGCCAGGGCTCCTCCCGGTTATCAATGTTCAAGTTGACATCCATCCATACCC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3377ss-RS/6K-GFP</td>
<td></td>
<td>ATGACAATTACTAAAAAATATTTCTAGGA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>psbA2-us-R2</td>
<td></td>
<td>ACTTTTTAAGTAAATTTGTATATTGTATATTGTATAATTTCTTATTTG</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene</th>
<th>Source</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFP-whole protein co-expression in Synechocystis</td>
<td>Gibson</td>
<td>G_futA1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G_futA2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P6</td>
</tr>
<tr>
<td>futA1-GFP</td>
<td>F</td>
<td>AGTCCATATGCTTCAAAAGTATCCCTGGTCG</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>CGAAAGCGCTCTTCTCAACCTGTTG</td>
</tr>
<tr>
<td>futA2-GFP</td>
<td>F</td>
<td>AGTCCATAGCAACAATGAACTTCCCAGCGCC</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>CGAAAGCGCTCTTCTCAACCTGTTG</td>
</tr>
<tr>
<td>GFP</td>
<td>F</td>
<td>AGTCCATATGCTGTAAGGGCCGAAGAGGCTTCC</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>CGAAAGCGCTCTTCTCAACCTGTTG</td>
</tr>
<tr>
<td>Tery_3377</td>
<td>pet_21a</td>
<td>3377_OE</td>
</tr>
<tr>
<td>Overexpression in E.Coli</td>
<td></td>
<td>R</td>
</tr>
</tbody>
</table>
Table A2: The 50 top differentially expressed genes in the transcriptomic analysis. Presented are each treatment’s average normalised read counts of the 50 most significantly differentially expressed genes in descending significance (P-value).

<table>
<thead>
<tr>
<th>Entry no.</th>
<th>Annotation</th>
<th>[Dust]</th>
<th>Dust+</th>
<th>Fe-</th>
<th>Fe+</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2561</td>
<td>cytochrome c, class I</td>
<td>18</td>
<td>195</td>
<td>30</td>
<td>608</td>
<td>1.1E-118</td>
</tr>
<tr>
<td>3345</td>
<td>hypothetical protein</td>
<td>176</td>
<td>106</td>
<td>263</td>
<td>1831</td>
<td>1.9E-51</td>
</tr>
<tr>
<td>2563</td>
<td>plastocyanin</td>
<td>2875</td>
<td>1654</td>
<td>2567</td>
<td>664</td>
<td>1.8E-48</td>
</tr>
<tr>
<td>0419</td>
<td>hemolysin-type calcium-binding protein</td>
<td>26</td>
<td>266</td>
<td>63</td>
<td>374</td>
<td>1.2E-39</td>
</tr>
<tr>
<td>2355</td>
<td>hemolysin-type calcium-binding protein</td>
<td>524</td>
<td>212</td>
<td>664</td>
<td>114</td>
<td>4.3E-38</td>
</tr>
<tr>
<td>0424</td>
<td>hemolysin-type calcium-binding protein</td>
<td>13</td>
<td>127</td>
<td>23</td>
<td>228</td>
<td>4.9E-37</td>
</tr>
<tr>
<td>0450</td>
<td>hypothetical protein</td>
<td>22</td>
<td>216</td>
<td>9</td>
<td>122</td>
<td>1.0E-33</td>
</tr>
<tr>
<td>3167</td>
<td>hemolysin-type calcium-binding protein</td>
<td>158</td>
<td>148</td>
<td>263</td>
<td>603</td>
<td>2.5E-28</td>
</tr>
<tr>
<td>1500</td>
<td>hypothetical protein</td>
<td>32</td>
<td>35</td>
<td>25</td>
<td>160</td>
<td>2.1E-25</td>
</tr>
<tr>
<td>0345</td>
<td>TENA/THI-4 protein</td>
<td>270</td>
<td>1514</td>
<td>132</td>
<td>677</td>
<td>1.1E-22</td>
</tr>
<tr>
<td>3559</td>
<td>hypothetical protein</td>
<td>64</td>
<td>180</td>
<td>143</td>
<td>337</td>
<td>5.0E-22</td>
</tr>
<tr>
<td>3376</td>
<td>phosphoglycerate kinase</td>
<td>230</td>
<td>475</td>
<td>372</td>
<td>706</td>
<td>1.3E-21</td>
</tr>
<tr>
<td>0346</td>
<td>hypothetical protein</td>
<td>4</td>
<td>57</td>
<td>1</td>
<td>38</td>
<td>2.5E-21</td>
</tr>
<tr>
<td>4114</td>
<td>Mo-dependent nitrogenase-like</td>
<td>651</td>
<td>342</td>
<td>702</td>
<td>272</td>
<td>5.2E-21</td>
</tr>
<tr>
<td>4763</td>
<td>photosystem q(b) protein</td>
<td>88774</td>
<td>80535</td>
<td>82336</td>
<td>37588</td>
<td>8.8E-21</td>
</tr>
<tr>
<td>0349</td>
<td>cell surface protein</td>
<td>40</td>
<td>542</td>
<td>5</td>
<td>229</td>
<td>2.0E-20</td>
</tr>
<tr>
<td>2376</td>
<td>caffeoyl-CoA O-methyltransferase</td>
<td>64</td>
<td>170</td>
<td>71</td>
<td>262</td>
<td>2.0E-20</td>
</tr>
<tr>
<td>0347</td>
<td>S-methyltetrahydropteroylglutamate/homocysteine-S-methyltransferase</td>
<td>99</td>
<td>2579</td>
<td>31</td>
<td>1347</td>
<td>1.1E-19</td>
</tr>
<tr>
<td>1368</td>
<td>ferredoxin-nitrite reductase</td>
<td>78</td>
<td>16</td>
<td>163</td>
<td>39</td>
<td>6.3E-19</td>
</tr>
<tr>
<td>1587</td>
<td>fructose-bisphosphate aldolase</td>
<td>24590</td>
<td>24346</td>
<td>28958</td>
<td>10777</td>
<td>1.6E-18</td>
</tr>
<tr>
<td>2360</td>
<td>hypothetical protein</td>
<td>46</td>
<td>37</td>
<td>51</td>
<td>144</td>
<td>2.2E-18</td>
</tr>
<tr>
<td>protein_id</td>
<td>protein</td>
<td>MW</td>
<td>R</td>
<td>kD</td>
<td>n</td>
<td>e-value</td>
</tr>
<tr>
<td>------------</td>
<td>-------------------------------------------------------------------------</td>
<td>----</td>
<td>---</td>
<td>----</td>
<td>---</td>
<td>---------</td>
</tr>
<tr>
<td>1047</td>
<td>AAA ATPase</td>
<td>8491</td>
<td>6707</td>
<td>11000</td>
<td>5830</td>
<td>2.4E-18</td>
</tr>
<tr>
<td>1070</td>
<td>assimilatory nitrate reductase</td>
<td>51</td>
<td>19</td>
<td>132</td>
<td>30</td>
<td>3.1E-18</td>
</tr>
<tr>
<td>0803</td>
<td>hypothetical protein</td>
<td>199</td>
<td>328</td>
<td>283</td>
<td>574</td>
<td>3.4E-18</td>
</tr>
<tr>
<td>2500</td>
<td>hypothetical protein</td>
<td>10</td>
<td>16</td>
<td>17</td>
<td>77</td>
<td>5.4E-18</td>
</tr>
<tr>
<td>2501</td>
<td>BadF/BadG/BcrA/BcrD type ATPase</td>
<td>108</td>
<td>260</td>
<td>112</td>
<td>642</td>
<td>1.3E-17</td>
</tr>
<tr>
<td>0844</td>
<td>putative iron-sulfur cluster-binding protein</td>
<td>25</td>
<td>158</td>
<td>23</td>
<td>87</td>
<td>4.7E-17</td>
</tr>
<tr>
<td>0685</td>
<td>OpcA protein</td>
<td>747</td>
<td>433</td>
<td>1259</td>
<td>566</td>
<td>1.4E-16</td>
</tr>
<tr>
<td>4209</td>
<td>phosphoenolpyruvate synthase</td>
<td>23</td>
<td>42</td>
<td>35</td>
<td>143</td>
<td>4.6E-16</td>
</tr>
<tr>
<td>2710</td>
<td>hemolysin-type calcium-binding protein</td>
<td>348</td>
<td>405</td>
<td>474</td>
<td>804</td>
<td>5.0E-16</td>
</tr>
<tr>
<td>1380</td>
<td>triosephosphate isomerase</td>
<td>289</td>
<td>448</td>
<td>210</td>
<td>176</td>
<td>1.9E-15</td>
</tr>
<tr>
<td>2902</td>
<td>two component transcriptional regulator</td>
<td>79</td>
<td>42</td>
<td>88</td>
<td>229</td>
<td>2.6E-15</td>
</tr>
<tr>
<td>2437</td>
<td>ATPase</td>
<td>348</td>
<td>451</td>
<td>753</td>
<td>763</td>
<td>5.7E-15</td>
</tr>
<tr>
<td>1667</td>
<td>photosystem antenna protein-like protein</td>
<td>18035</td>
<td>15593</td>
<td>17542</td>
<td>7566</td>
<td>7.9E-15</td>
</tr>
<tr>
<td>1048</td>
<td>hypothetical protein</td>
<td>6858</td>
<td>4970</td>
<td>7448</td>
<td>3866</td>
<td>1.1E-14</td>
</tr>
<tr>
<td>0677</td>
<td>glutaredoxin-like protein</td>
<td>650</td>
<td>440</td>
<td>578</td>
<td>313</td>
<td>2.3E-14</td>
</tr>
<tr>
<td>1355</td>
<td>hemolysin-type calcium-binding protein</td>
<td>277</td>
<td>435</td>
<td>409</td>
<td>722</td>
<td>4.1E-14</td>
</tr>
<tr>
<td>0672</td>
<td>hypothetical protein</td>
<td>340</td>
<td>205</td>
<td>344</td>
<td>120</td>
<td>6.7E-14</td>
</tr>
<tr>
<td>0843</td>
<td>hypothetical protein</td>
<td>8</td>
<td>79</td>
<td>8</td>
<td>42</td>
<td>2.2E-13</td>
</tr>
<tr>
<td>4280</td>
<td>hypothetical protein</td>
<td>3771</td>
<td>2474</td>
<td>3960</td>
<td>2205</td>
<td>8.4E-13</td>
</tr>
<tr>
<td>4617</td>
<td>cobalamin synthesis protein, P47K</td>
<td>2992</td>
<td>1623</td>
<td>1810</td>
<td>1006</td>
<td>1.3E-12</td>
</tr>
<tr>
<td>3320</td>
<td>acyl-[acyl-carrier-protein]–UDP-N-acetylglucosamine O-acyltransferase</td>
<td>13</td>
<td>26</td>
<td>31</td>
<td>80</td>
<td>1.3E-12</td>
</tr>
<tr>
<td>0489</td>
<td>undecaprenyl-phosphate galactosephosphotransferase</td>
<td>184</td>
<td>281</td>
<td>106</td>
<td>161</td>
<td>2.1E-12</td>
</tr>
<tr>
<td>1230</td>
<td>photosystem II D2 protein (photosystem q(a) protein)</td>
<td>18551</td>
<td>18196</td>
<td>17655</td>
<td>9876</td>
<td>2.2E-12</td>
</tr>
<tr>
<td>2486</td>
<td>phycobilisome linker polypeptide</td>
<td>1099</td>
<td>750</td>
<td>1135</td>
<td>445</td>
<td>2.2E-12</td>
</tr>
</tbody>
</table>
Table A3: The 50 most highly expressed genes in the transcriptomic analysis. Expression intensity (normalised read counts) is calculated as the average of all treatments and is presented in descending order for the 50 most highly expressed genes. Significantly differentially expressed genes are indicated (bold).

<table>
<thead>
<tr>
<th>Tery no.</th>
<th>Annotation</th>
<th>[Dust]</th>
<th>Dust+</th>
<th>Fe-</th>
<th>Fe+</th>
<th>Average</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>4616</td>
<td>GTP cyclohydrolase I</td>
<td>4047</td>
<td>2966</td>
<td>3129</td>
<td>1574</td>
<td>4.1E-12</td>
<td></td>
</tr>
<tr>
<td>1880</td>
<td>peptidoglycan binding domain-containing protein</td>
<td>298</td>
<td>472</td>
<td>253</td>
<td>644</td>
<td>8.1E-12</td>
<td></td>
</tr>
<tr>
<td>4355</td>
<td>cysteine desulphurase activator complex subunit SufB</td>
<td>449</td>
<td>328</td>
<td>640</td>
<td>304</td>
<td>1.0E-11</td>
<td></td>
</tr>
<tr>
<td>1068</td>
<td>glucose-6-phosphate 1-dehydrogenase</td>
<td>1114</td>
<td>626</td>
<td>1782</td>
<td>790</td>
<td>1.8E-11</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tery no.</th>
<th>Annotation</th>
<th>[Dust]</th>
<th>Dust+</th>
<th>Fe-</th>
<th>Fe+</th>
<th>Average</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2820</td>
<td>hypothetical protein</td>
<td>947</td>
<td>439</td>
<td>962</td>
<td>201</td>
<td>1.9E-11</td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>Description</td>
<td>Value 1</td>
<td>Value 2</td>
<td>Value 3</td>
<td>Value 4</td>
<td>Value 5</td>
<td>Value 6</td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>1666</td>
<td>flavodoxin FlmA</td>
<td>7753</td>
<td>9020</td>
<td>9475</td>
<td>6166</td>
<td>8164</td>
<td>9E-03</td>
</tr>
<tr>
<td>1047</td>
<td>AAA ATPase</td>
<td>8491</td>
<td>6707</td>
<td>11000</td>
<td>5830</td>
<td>8067</td>
<td>2E-18</td>
</tr>
<tr>
<td>3826</td>
<td>dimethylaniline monoxygenase</td>
<td>9681</td>
<td>8151</td>
<td>8955</td>
<td>4944</td>
<td>7933</td>
<td>4E-06</td>
</tr>
<tr>
<td>2484</td>
<td>photosystem antenna protein-like protein</td>
<td>7079</td>
<td>8319</td>
<td>9810</td>
<td>5956</td>
<td>7791</td>
<td>6E-03</td>
</tr>
<tr>
<td>2324</td>
<td>gas vesicle synthesis protein GvpA</td>
<td>9545</td>
<td>6622</td>
<td>7223</td>
<td>6208</td>
<td>7399</td>
<td>2E-01</td>
</tr>
<tr>
<td>4669</td>
<td>photosystem I P700 chlorophyll a apoprotein A1</td>
<td>4496</td>
<td>8663</td>
<td>4401</td>
<td>7450</td>
<td>6252</td>
<td>9E-07</td>
</tr>
<tr>
<td>4367</td>
<td>hypothetical protein</td>
<td>7409</td>
<td>5057</td>
<td>7104</td>
<td>3578</td>
<td>5787</td>
<td>4E-09</td>
</tr>
<tr>
<td>1048</td>
<td>hypothetical protein</td>
<td>6858</td>
<td>4970</td>
<td>7448</td>
<td>3866</td>
<td>5785</td>
<td>1E-14</td>
</tr>
<tr>
<td>3372</td>
<td>beta-lg H3/fasciclin</td>
<td>8555</td>
<td>4009</td>
<td>7139</td>
<td>3023</td>
<td>5682</td>
<td>1E-07</td>
</tr>
<tr>
<td>4136</td>
<td>nitrogenase iron protein subunit NifH</td>
<td>3315</td>
<td>5025</td>
<td>5708</td>
<td>7087</td>
<td>5284</td>
<td>3E-02</td>
</tr>
<tr>
<td>4099</td>
<td>fructose-1,6-bisphosphate aldolase</td>
<td>4804</td>
<td>5856</td>
<td>5567</td>
<td>4436</td>
<td>5166</td>
<td>6E-02</td>
</tr>
<tr>
<td>4137</td>
<td>nitrogenase molybdenum-iron protein subunit alpha</td>
<td>3019</td>
<td>4567</td>
<td>5781</td>
<td>6661</td>
<td>50C7</td>
<td>2E-01</td>
</tr>
<tr>
<td>4410</td>
<td>ribulose bisphosphate carboxylase</td>
<td>3296</td>
<td>7294</td>
<td>4267</td>
<td>4835</td>
<td>4923</td>
<td>8E-05</td>
</tr>
<tr>
<td>4668</td>
<td>photosystem I P700 chlorophyll a apoprotein A2</td>
<td>3410</td>
<td>7244</td>
<td>3295</td>
<td>5696</td>
<td>4912</td>
<td>4E-08</td>
</tr>
<tr>
<td>0397</td>
<td>phage tail Collar</td>
<td>4017</td>
<td>6691</td>
<td>4330</td>
<td>4445</td>
<td>4817</td>
<td>2E-02</td>
</tr>
<tr>
<td>0183</td>
<td>photosystem q(b) protein</td>
<td>6927</td>
<td>4833</td>
<td>4437</td>
<td>2066</td>
<td>4566</td>
<td>3E-07</td>
</tr>
<tr>
<td>2559</td>
<td>flavodoxin</td>
<td>5124</td>
<td>3238</td>
<td>5397</td>
<td>3482</td>
<td>4310</td>
<td>8E-05</td>
</tr>
<tr>
<td>1776</td>
<td>type 12 methyltransferase</td>
<td>2622</td>
<td>4036</td>
<td>2867</td>
<td>4685</td>
<td>3552</td>
<td>1E-03</td>
</tr>
<tr>
<td>4138</td>
<td>nitrogenase molybdenum-iron protein subunit beta</td>
<td>2317</td>
<td>2986</td>
<td>3955</td>
<td>4288</td>
<td>3386</td>
<td>6E-02</td>
</tr>
<tr>
<td>0600</td>
<td>hypothetical protein</td>
<td>2226</td>
<td>2913</td>
<td>3982</td>
<td>4352</td>
<td>3368</td>
<td>3E-01</td>
</tr>
<tr>
<td>1137</td>
<td>cytochrome b6</td>
<td>3276</td>
<td>3294</td>
<td>3411</td>
<td>2676</td>
<td>3275</td>
<td>3E-02</td>
</tr>
<tr>
<td>1798</td>
<td>apocytochrome f</td>
<td>3456</td>
<td>4218</td>
<td>2754</td>
<td>2570</td>
<td>3250</td>
<td>2E-06</td>
</tr>
<tr>
<td>2204</td>
<td>FO1 ATP synthase subunit A</td>
<td>2668</td>
<td>4335</td>
<td>2781</td>
<td>3182</td>
<td>3241</td>
<td>8E-06</td>
</tr>
<tr>
<td>3819</td>
<td>beta-ketoacyl synthase</td>
<td>2862</td>
<td>3197</td>
<td>3620</td>
<td>3132</td>
<td>32C3</td>
<td>4E-01</td>
</tr>
<tr>
<td>0997</td>
<td>phycobilisome protein</td>
<td>2313</td>
<td>5254</td>
<td>2359</td>
<td>2654</td>
<td>3145</td>
<td>1E-05</td>
</tr>
<tr>
<td></td>
<td>Protein Name</td>
<td>Value1</td>
<td>Value2</td>
<td>Value3</td>
<td>Value4</td>
<td>Value5</td>
<td>Value6</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>4280</td>
<td>hypothetical protein</td>
<td>3771</td>
<td>2474</td>
<td>3960</td>
<td>2205</td>
<td>3103</td>
<td>8E-13</td>
</tr>
<tr>
<td>4661</td>
<td>S-adenosylmethionine synthetase</td>
<td>2374</td>
<td>3796</td>
<td>2465</td>
<td>3562</td>
<td>3049</td>
<td>1E-05</td>
</tr>
<tr>
<td>5048</td>
<td>phycocyanin subunit alpha</td>
<td>2768</td>
<td>4528</td>
<td>2373</td>
<td>2293</td>
<td>2991</td>
<td>2E-06</td>
</tr>
<tr>
<td>4327</td>
<td>chaperonin GroEL</td>
<td>2068</td>
<td>3801</td>
<td>3174</td>
<td>2871</td>
<td>2979</td>
<td>1E-04</td>
</tr>
<tr>
<td>4616</td>
<td>GTP cyclohydrolase I</td>
<td>4047</td>
<td>2966</td>
<td>3129</td>
<td>1574</td>
<td>2929</td>
<td>4E-12</td>
</tr>
<tr>
<td>0996</td>
<td>phycobilisome protein</td>
<td>2125</td>
<td>5052</td>
<td>1945</td>
<td>2278</td>
<td>2850</td>
<td>4E-06</td>
</tr>
<tr>
<td>1834</td>
<td>6-phosphogluconate dehydrogenase</td>
<td>2349</td>
<td>2099</td>
<td>4345</td>
<td>2500</td>
<td>2823</td>
<td>2E-05</td>
</tr>
<tr>
<td>0998</td>
<td>phycobilisome protein</td>
<td>2272</td>
<td>4688</td>
<td>1950</td>
<td>2245</td>
<td>2789</td>
<td>1E-06</td>
</tr>
<tr>
<td>2900</td>
<td>hypothetical protein</td>
<td>2846</td>
<td>1844</td>
<td>3973</td>
<td>2452</td>
<td>2779</td>
<td>9E-07</td>
</tr>
<tr>
<td>5049</td>
<td>phycocyanin subunit beta</td>
<td>2390</td>
<td>4118</td>
<td>2370</td>
<td>2226</td>
<td>2776</td>
<td>6E-05</td>
</tr>
<tr>
<td>0117</td>
<td>peptidase S8/S53 subtilisin kexin sedolisin</td>
<td>1977</td>
<td>2660</td>
<td>2789</td>
<td>3538</td>
<td>2741</td>
<td>1E-06</td>
</tr>
<tr>
<td>4030</td>
<td>glyceraldehyde-3-phosphate dehydrogenase</td>
<td>2345</td>
<td>3114</td>
<td>2567</td>
<td>2801</td>
<td>2707</td>
<td>2E-02</td>
</tr>
</tbody>
</table>
Figure A.1: Clustered expression patterns across the four experimental treatments. Pearson product moment correlation coefficient between each pair of differentially expressed genes and WPGMA was used to cluster expression profiles across the three replicate cultures of each of the 4 experimental treatments: [Dust], Dust+, Fe- and Fe+, in 10 groups.


employing iron (II) as a cofactor to bind the operator of an iron transport operon in *Escherichia coli*. Biochemistry, 26(17), 5471–7.
Bibby, T., Nield, J., & Barber, J. (2001). Iron deficiency induces the formation of an antenna


Falkowski, P. G., Katz, M. E., Knoll, A. H., Quigg, A., Raven, J. a, Schofield, O., & Taylor, F. J. R.


Bibliography


Hewson, I., Govil, S., Capone, D., Carpenter, E., & Fuhrman, J. (2004). Evidence of *Trichodesmium* viral lysis and potential significance for biogeochemical cycling in the...


Bibliography


Bibliography

347(6226), 1066–1067.
Ludwig, M., & Bryant, D. A. (2011). Transcription Profiling of the model cyanobacterium Synechococcus sp. Strain PCC 7002 by Next-Gen (SOLiDTM) sequencing of cDNA. Frontiers in Microbiology, 2, 41.1


Bibliography


---

213


Tuit, C., Waterbury, J., & Ravizza, G. (2004). Diel variation of molybdenum and iron in...


Trichodesmium spp. Marine Ecology Progress Series, 323, 35–45.


