

## ARTICLE OPEN



# Transcriptional responses of *Trichodesmium* to natural inverse gradients of Fe and P availability

E. Cerdan-Garcia<sup>1</sup> , A. Baylay<sup>1</sup>, D. Polyviou<sup>2</sup>, E. M. S. Woodward<sup>1</sup> , L. Wrightson<sup>4</sup>, C. Mahaffey<sup>4</sup>, M. C. Lohan<sup>1</sup> , C. M. Moore<sup>1</sup>, T. S. Bibby<sup>1</sup> and J. C. Robidart<sup>2</sup> 

© The Author(s) 2021

The filamentous diazotrophic cyanobacterium *Trichodesmium* is responsible for a significant fraction of marine di-nitrogen (N<sub>2</sub>) fixation. Growth and distribution of *Trichodesmium* and other diazotrophs in the vast oligotrophic subtropical gyres is influenced by iron (Fe) and phosphorus (P) availability, while reciprocally influencing the biogeochemistry of these nutrients. Here we use observations across natural inverse gradients in Fe and P in the North Atlantic subtropical gyre (NASG) to demonstrate how *Trichodesmium* acclimates in situ to resource availability. Transcriptomic analysis identified progressive upregulation of known iron-stress biomarker genes with decreasing Fe availability, and progressive upregulation of genes involved in the acquisition of diverse P sources with decreasing P availability, while genes involved in N<sub>2</sub> fixation were upregulated at the intersection under moderate Fe and P availability. Enhanced N<sub>2</sub> fixation within the Fe and P co-stressed transition region was also associated with a distinct, consistent metabolic profile, including the expression of alternative photosynthetic pathways that potentially facilitate ATP generation required for N<sub>2</sub> fixation with reduced net oxygen production. The observed response of *Trichodesmium* to availability of both Fe and P supports suggestions that these biogeochemically significant organisms employ unique molecular, and thus physiological responses as adaptations to specifically exploit the Fe and P co-limited niche they construct.

The ISME Journal; <https://doi.org/10.1038/s41396-021-01151-1>

## INTRODUCTION

In low nitrogen (N) marine environments, diazotrophs, organisms capable of fixing atmospheric di-N into biologically available ammonia (NH<sub>3</sub>) [1, 2] contribute to a significant fraction of new N input in the oceans. The colonial cyanobacterium *Trichodesmium* is one of the main contributors to the fixed N budget of the world's oceans [3, 4] accounting for up to 50% of total upper ocean N<sub>2</sub> fixation in some areas [5] with an estimated annual input of 60–80 Tg N [6]. *Trichodesmium* can be the dominant diazotroph in regions such as the North Atlantic subtropical gyre (NASG), fuelling biological productivity and driving biogeochemical cycles [5–7].

*Trichodesmium*'s distribution and N<sub>2</sub> fixation activity is restricted by the availability of nutrients such as iron (Fe) and phosphorus (P) [7–9]. This is distinctly evident across latitudinal gradients in the Atlantic, where low dust Fe flux south of the inter-tropical convergence zone (ITCZ) in the South Atlantic subtropical gyre limits the activity of *Trichodesmium* and other N<sub>2</sub> fixers, resulting in relatively high residual P concentrations [9, 10]. In contrast, increased dust flux north of the ITCZ provides sufficient Fe for *Trichodesmium* to flourish and thus drawdown surface P to nanomolar concentrations [7, 11–13]. Such boundaries between high-Fe/low-P and low-Fe/high-P regions thus represent constructed niches which appear widespread in low latitude oligotrophic oceans [9, 11, 14]. Moreover, evidence for Fe and P co-limitation [15] and observed physiological responses of *Trichodesmium* under these conditions have led to suggestions

that the organism may be specifically adapted to nutrient co-stress [16–19].

Culture and field studies of *Trichodesmium* using gene expression analysis [20], proteomic profiling [19, 21] or both [18, 22] have provided insights into *Trichodesmium*'s ecophysiology, including the identification of biomarker indicators of nutrient stress [17, 20, 23]. Given the importance of Fe and P availability, *Trichodesmium* has evolved several acclimation strategies to cope when these key resources are restricted. This is particularly true for Fe, given its absolute requirement in both the catalysis of N<sub>2</sub> fixation and photosynthesis [24, 25]. *Trichodesmium* does not temporally separate photosynthesis from N<sub>2</sub> fixation and as such cannot 'share' Fe between these processes, potentially increasing Fe requirements relative to other diazotrophic groups such as *Crocospaera* [16, 21]. In response to Fe scarcity, *Trichodesmium* reduces demand through reduction of high Fe content enzymes including photosystems I and II (PSI and PSII) and nitrogenase [21, 24] and replacement of Fe-containing enzymes with non-Fe dependent isozymes such as ferredoxin with flavodoxin [26] and cytochrome c<sub>533</sub> with the Cu-dependent plastocyanin [27]. Moreover, *Trichodesmium* potentially increases Fe supply through upregulation of Fe transporters (*IdiA/FutA*) [28, 29]. Simultaneously performing photosynthesis and N<sub>2</sub> fixation generates a further challenge for *Trichodesmium* in generating the ATP and reductant required for N<sub>2</sub> fixation while minimising net oxygen production to protect nitrogenase from

<sup>1</sup>Ocean and Earth Science, University of Southampton, Southampton SO14 3ZH, UK. <sup>2</sup>National Oceanography Centre, Southampton SO14 3ZH, UK. <sup>3</sup>Plymouth Marine Laboratory, Plymouth PL1 3DH, UK. <sup>4</sup>Earth, Ocean and Ecological Sciences, University of Liverpool, Liverpool L69 3BX, UK. ✉email: e.cerdangarcia@soton.ac.uk; j.robidart@noc.ac.uk

Received: 15 April 2021 Revised: 28 October 2021 Accepted: 1 November 2021

Published online: 24 November 2021

oxygen inhibition [30]. While regulation over the light cycle [31] and reversible cellular differentiation into proposed diazocytes [4, 32] can help achieve this, the molecular regulation that enables simultaneous aerobic and anaerobic chemistry is not fully characterised.

*Trichodesmium* can access P in various forms [33]. In addition to acquisition of dissolved inorganic phosphate (DIP) (including phosphate and phosphite), *Trichodesmium* can utilise a broad range of dissolved organic P (DOP) compounds when DIP is low, which is used as an indication of P stress state. These compounds include phosphomonoesters and phosphonates [34], which can be found at higher concentrations in some oligotrophic regions [35]. A range of P-stress genes and protein biomarkers involved in P-acquisition have thus been characterised, such as the high-affinity inorganic P transporters *pstS* and *sphX*, the putative alkaline phosphatases (APs) *phoA* and *phoX* [22, 33], the phosphonate-related *phnCDEEGHIJKLM* [34], and the phosphite utilisation *ptxABCD* genes [36].

Here, *Trichodesmium*'s transcriptomic response and relative abundance within the diazotrophic community was investigated across naturally opposed gradients in Fe and P concentrations. A reduced capability for  $N_2$  fixation is observed at both ends of the transect, where either Fe or P are at their lowest concentrations. Elevated  $N_2$  fixation rates in the transition region between the areas of lowest P and lowest Fe availability coincided with a consistent gene expression profile, suggesting an increased capacity to fix  $N_2$  and produce cobalamin ( $B_{12}$  vitamin) at intermediate nutrient concentrations. In addition, the enhanced  $N_2$  fixation capacity may be supported by alternative ATP generation pathways that are both Fe-efficient and do not generate net  $O_2$ . The observed functional acclimatisation dependent on Fe and P availability may be key contributing factor in the ecological success of *Trichodesmium* across oceanic gyres [14].

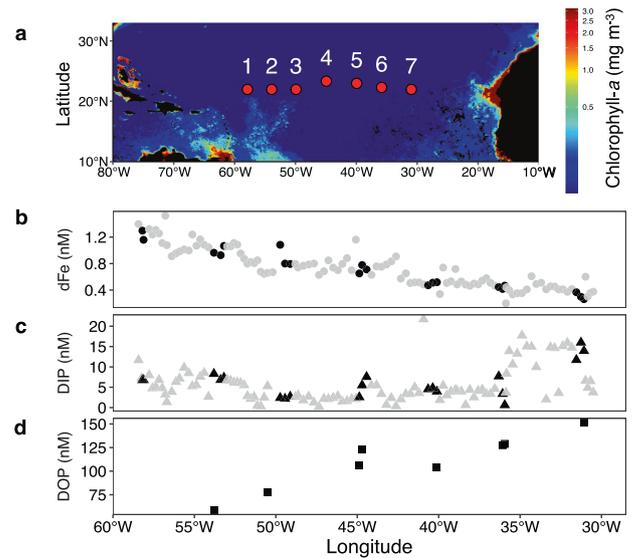
## MATERIALS AND METHODS

### Hydrography and environmental data

Sampling was carried out on board the RRS James Cook JC150 (GEOTRACES GApr08) in the subtropical North Atlantic on a transect from Guadeloupe (French Caribbean) to Tenerife (Canary Islands) (26th June to 12th August 2017) (Fig. 1a). Surface (2–3 m) underway water sampling for nutrient and Fe analysis was conducted using a Teflon diaphragm pump (Almatec A-15) connected to a "Towed-Fish", which pumped seawater into a class-1000 clean-air laboratory. Samples for dissolved Fe (dFe) were filtered in-line (<0.2  $\mu$ m; Sartobran) and measured using flow injection with chemiluminescence detection [37] as reported in Kunde et al. [38]. The detection limits were  $0.03 \pm 0.02$  nM ( $n = 59$ ). Nanomolar measurements of DIP were made using colorimetric method with segmented flow analysis coupled to a 2 m liquid waveguide as the analytical flow cell to improve detection limits [39]. DIP detection limit was 1 nM ( $3 \times$  the Milli-Q water baseline, determined over the course of the research cruise). Certified reference materials from KANSO Technos (Japan) were measured for quality control. Total dissolved phosphorus (TDP) was determined using the high temperature acid persulfate technique as in Davis et al. [40]. DOP was taken as the difference between TDP and DIP (DOP = TDP-DIP, where DOP detection limit = 40 nM; TDP-DIP, thus  $2 \times$  DL of DIP (20 nM)) [41]. In situ  $N_2$  fixation rates were measured using the modified bubble injection method described by Klawonn et al. [42]. Specifically, 8 mL of 98%  $^{15}N_2$  gas (Cambridge Isotope Laboratories, NLM-363-0, lot No: I-21065/AR0664758) was added to each 4.2 L incubation bottle through the septa. Bottles were strapped to a cable drum, rolled across the deck (~10 rpm) for 15 min. 20 mL of seawater was removed (4200 – 20 mL = 4180 mL, 0.47% volume) and filtered seawater was added to displace the volume removed, ensuring there was no bubble remaining in the bottle before the incubation (Supplementary information). Samples were analysed on board using a Hidden Membrane Inlet Mass Spectrometer (Supplementary information).

### Sample collection

*Trichodesmium* colonies were collected at seven stations along the transect (Fig. 1a) using plankton nets (200  $\mu$ m mesh, 50 cm diameter) (Duncan &



**Fig. 1** Transect and surface nutrient profiles. **a** Sampling stations 1–7 for the RRS James Cook-JC150 (GEOTRACES GApr08) cruise across the subtropical North Atlantic Ocean (June–August 2017 at ~22°N) from Guadeloupe (French Caribbean) to Tenerife (Canary Islands). Background colour shows the satellite-derived chlorophyll-*a* concentration ( $mg/m^3$ ) from July 2017 [100]. Surface nutrient concentrations (nM) for **(b)** dFe, **(c)** DIP and **(d)** DOP (in grey) including measurements from the seven sampling stations (in black). dFe is a sub-set of data from Kunde et al. [38].

Associates, UK) deployed at 15 meters for 15 min. Net cod-ends were emptied into an acid-cleaned bucket and transferred to a laminar flow hood inside a trace-metal clean bubble laboratory. Samples of 50 colonies (in triplicate) were randomly hand-picked using plastic Pasteur pipettes and individually rinsed through three steps (0.2  $\mu$ m filtered "Towed-Fish" seawater) before allocation into 150 mL bottles (previously HCl rinsed and filled with filtered <0.2  $\mu$ m "Towed-Fish" seawater). Bottles were filtered onto 0.2  $\mu$ m Supor filters (Sigma-Aldrich, St. Louis, USA), flash frozen in liquid  $N_2$  and stored at  $-80^\circ C$ . Samples were collected at the same time-point to remove the influence of any diel transcriptional variability. For this study we prioritised daybreak (5–6am local time) to measure transcription at peak nitrogenase, N metabolism, P metabolism and photosystem transcription according to Frischkorn et al. [43]. It is possible that this choice relegated transcripts from other gene families (troughs in diel transcription) to below the limit of detection. These include transcripts coding for amino acid metabolism, ATP synthesis, ribosomes and some cofactors and vitamins. In this manuscript we restrict our analyses to transcripts that are differentially expressed, and present in all samples across the transect.

### DNA/RNA extractions

RNA extractions used a modification of the AllPrep DNA/RNA mini kit (Qiagen, Manchester, UK), DNase digestion (Qiagen DNase kit) and clean-up (ZYMO RNA Clean and Concentrator) as described in Tang et al. [44]. RNA concentrations were measured using Qubit (Qiagen) and extracts were stored at  $-80^\circ C$  in 15  $\mu$ L aliquots for metatranscriptomic sequencing.

### Quantitative polymerase chain reaction (qPCR)

Targeted quantitative Polymerase Chain Reaction (qPCR) approaches were used to identify and quantify the *Trichodesmium* and UCYN-A *nifH* genes along the cruise transect (Supplementary Fig. 1). Gene copies were quantified relative to standard curves obtained from the amplification of linearised plasmids with the targeted gene inserts. Standard gene dilution series ( $10^7$ – $10^0$  copies  $L^{-1}$ ) were made with nuclease-free  $H_2O$  (Ambion, CA, USA) for each reaction and negative controls contained nuclease-free  $H_2O$  in place of DNA template (Methods described in Supplementary information).

## DNA/RNA sequencing

***nifH* amplicon sequencing.** Nested degenerate polymerase chain reaction (PCR) was used to amplify the *nifH* genes using degenerate primers [45] and second round PCR primers *nifH1* and *nifH2* [46] with Illumina tag sequences as described elsewhere [44, 47], in PCR reactions were cleaned and concentrated using the MinElute kit (Qiagen) and Nextera XT indexed PCR was used to add i5/i7 sequencing adaptors (Part 15044223 Rev. B, Illumina). Equimolar concentrations of *nifH* amplicons were pooled and sequenced on the MiSeq (Illumina) at the University of Southampton's Environmental Sequencing Facility.

**Whole transcriptome sequencing.** Depletion of rRNA from total RNA was performed using the Ribo-Zero Bacteria Magnetic Kit (Illumina). RNA-Seq libraries were prepared in triplicate for each sample for MiSeq (Illumina) sequencing. The library was prepared using the ScriptSeq v2 RNA-Seq Library Preparation Kit Illumina (originally EpiCentre), implementing a tagging oligo method [48] to determine which strand the transcript was from.

## Metagenome sequencing and assembly

Metagenome assembly came from a set of 56 single *Trichodesmium* colonies collected from 21st to 27th July 2016 aboard the RV Atlantic Explorer in the Sargasso Sea (NCBI BioProject ID PRJNA721834). Colonies were picked, filtered and frozen as described above (see section *Sample collection*). DNA was extracted with the DNeasy Plant mini kit (Qiagen) according to the manufacturer's protocol. DNA extracts from five *T. thiebautii* colonies were chosen for whole genome sequencing (based on DNA concentration). Library preparation used KAPA HyperPlus DNA library prep kit following manufacturer's instructions. Libraries were pooled and 2 × 300 bp paired-end sequencing was carried out on a MiSeq (Illumina) using a MiSeq v3 Reagent Kit (Illumina). Reads were trimmed using Cutadapt version 2.3 [49]. Initial assemblies were carried out using SPADes version 3.13.1 [50] in metagenome mode. Resulting contigs were binned using MaxBin [51], followed by CheckM [52]. All five samples yielded a distinct cyanobacterial bin estimated to be >97% complete based on marker gene analysis. To identify *Trichodesmium*-only reads, tetranucleotide frequency profiles for all metagenome contigs of length >5 kb were compared to those of randomly simulated fragments (500 fragments, mean length 10 kb, standard deviation 1 kb) from the IMS101 genome sequence (Supplementary Fig. 2). Original read sets were filtered to extract *Trichodesmium* reads by alignment to the binned contigs using Bowtie2 [53] and reassembled using SPADes. We picked assembly "sct-8t3" as a reference genome as it was the most contiguous.

Putative coding sequences were identified on the metagenome contigs using Prodigal version 2.6.3 [54] with default settings. The resulting 6669 predicted protein sequences were clustered into 4044 orthologous groups (OGs) using MCL version 14-137 [55] (inflation parameter 1.4) [43]. OGs were characterised by comparison against IMS101 protein sequences, and the SEED [56] and UniRef90 databases using DIAMOND [57]. Consensus annotations for each OG were chosen by taking the most abundant SEED and UniRef90 annotations across the proteins making up each OG, as has been previously described [23, 43]. Where different annotations were equally abundant within an OG, the hit with the highest bit score was chosen.

## Bioinformatics

**Amplicon sequencing.** *nifH* relative amplicon abundance was determined using QIIME [58] as in Tang et al. [44] Raw sequences were merged and quality filtered (quality score >20 and read length >200 bp) using USEARCH [59]. Following a chimera removal step, sequences were clustered into operational taxonomic units (OTUs) at 97% sequence similarity using UPARSE [60]. The *nifH* sequences obtained were assigned taxonomies using the BLAST resource on NCBI.

**Metatranscriptome dataset.** Raw sequence quality was assessed (FastQC version 11.9), before pre-processing, annotation and analysis using the SAMSA2 metatranscriptomic analysis pipeline [61]. Low-quality and short sequences and adaptors were trimmed with Trimmomatic v0.3 [62], followed by removal of RNA ribosomal sequences with SortMeRNA [63]. Paired-end reads were merged with PEAR v0.9.10 [64]. Cleaned sequences were annotated by DIAMOND sequence aligner (version 0.8.3) mapped against NCBI's most recent RefSeq non-redundant protein database (created January 2019) [57, 65], and against SEED [56] for hierarchical clustering of functional annotations. Mapping rates to *Trichodesmium*

averaged 36% across samples, similar to a previously published metatranscriptome (37%; [23, 43]). Reads were also mapped to contigs from the custom *T. thiebautii* metagenome from the Sargasso Sea (described above) using Hisat2 version 2.2.12 [66] in strand-specific mode. Reads mapping to predicted coding sequences were counted using featureCounts [67]. Using the Sargasso Sea reference, we obtained higher mapping rates (41.92%; Supplementary Table 1), which could be due to the inclusion of genes from other *Trichodesmium* species in the field in addition to *T. erythraeum* IMS101, as explained in Rouco et al. [23] The broad data interpretations were similar when using the more limited IMS101 mapping. Therefore, the custom metagenome was used for subsequent analysis.

## Differential expression and statistical analyses

Differential gene expression analysis was carried out using the R (4.0.3) package DESeq2 (1.28.1) [20, 68, 69]. DESeq2 only considers genes that fit into a binomial distribution and are differentially expressed between stations. The threshold to include transcripts into the analysis depends on sample size and the number of parameters to be estimated. OG relative abundances were estimated from the package negative binomial generalised linear model fit for each OG. Significance was calculated through pairwise comparisons between stations by a likelihood ratio test with an adjusted *p* value (<0.1) for multiple test correction, similar to Walworth et al. [17] Differential expression was calculated relative to the Rota housekeeping OG [23, 33].

Normal distribution of data was based on the total library size using the median ratio method implemented in DESeq2 [68]. Pearson parametric tests (*p* value level of significance 0.05) were used for the correlations between the normalised gene expression across all samples (*n* = 17) and measured environmental conditions.

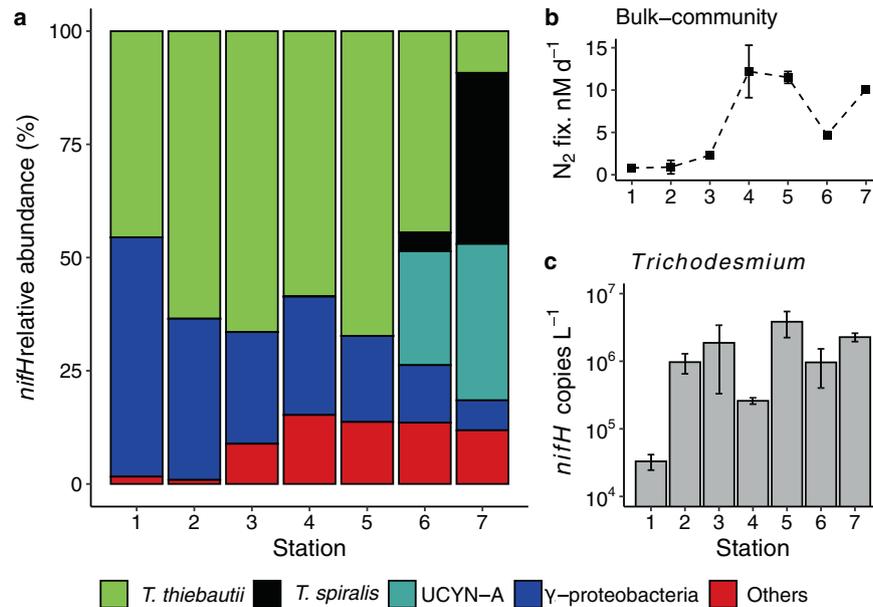
## RESULTS AND DISCUSSION

### Environmental and hydrographic data

Seven stations were sampled along a west-east transect across the NASG at ~22°N (Fig. 1a). A west to east decrease of surface total dissolved iron (dFe) concentrations was observed, decreasing from 1.23 to 0.26 nM (Fig. 1b), consistent with other oligotrophic North Atlantic studies [70, 71] and reflecting the influence of advection in the west [38]. Opposing the dFe gradient, an increasing eastward trend was observed for DOP and total dissolved surface P (TDP), while DIP was slightly elevated at the easternmost stations (6 and 7) (Fig. 1c, d). The low concentrations of surface DIP (0.6–15 nM) and DOP (72–150 nM) were consistent with conditions typically observed in the North Atlantic [7, 12, 72]. The dFe:TDP ratio thus decreased from west ( $12.3 \pm 2.6$  mmol:mol) to east ( $2.0 \pm 0.2$  mmol:mol) (Supplementary Fig. 3), stations 3, 4 and 5 (mid-transect) being located within the transition zone between the higher relative concentrations of dFe and TDP. Notably, there was no zonal trend in DIN (Supplementary Table 2).

### Diazotroph diversity and *Trichodesmium* abundance

Diazotroph community composition was revealed by *nifH* gene (which encodes a core subunit of the nitrogenase enzyme responsible for N<sub>2</sub> fixation) ([73] Fig. 2a) sequencing. In addition to *Trichodesmium*, a number of microorganisms actively fix atmospheric N<sub>2</sub> in oligotrophic systems [74]. Total *nifH* sequence diversity increased from west to east, and *Trichodesmium* spp. sequences dominated throughout, averaging ~57% of total *nifH* amplicon sequences (ranging between 46 and 67%), similar to other studies from the region [75]. *T. thiebautii* was identified as the dominant *Trichodesmium* species at stations 1–6, whereas *T. spiralis* was the dominant species at the easternmost station 7, where dFe:TDP was lowest. Despite their morphological differences, both *Trichodesmium* species belong to the same phylogenetic clade (Clade I) showing different biogeography [76]. The widespread and abundant gamma-proteobacteria group  $\gamma$ -24774A11 [77, 78] was a significant part of the diazotroph community, averaging 25% of total *nifH* DNA sequences, gradually decreasing in relative abundance from west to east (57–7%).



**Fig. 2** Diazotroph abundances and measured bulk  $N_2$  fixation rates. **a** Relative abundance of the *nifH* amplicon for *Trichodesmium* sp. (*T. thiebautii*, *T. spiralis*), UCYN-A,  $\gamma$ -Proteobacterium 24774A11 and other non-identified *nifH* sequences. **b** In situ bulk measured  $N_2$  fixation rates (nM per day). **c** *Trichodesmium nifH* gene abundance (genes  $L^{-1}$ ) quantified by qPCR.

Lastly, the unicellular UCYN-A was only detected at the easternmost stations (st. 6,7) representing 25% and 34% of total *nifH* DNA sequences, respectively, potentially suggesting UCYN-A is competitive under lower dFe:TDP, and gamma-proteobacteria under higher dFe:TDP. A strong positive relationship between UCYN-A and phosphate concentration has been observed previously [79] including in the North Atlantic [44], though this relationship doesn't hold for all regions [80]. While the gamma-proteobacteria have been found to be more abundant at higher temperatures [81, 82], their relationship with nutrients is not as clear. In situ community  $N_2$  fixation rates ranged from 0.8 to 18.2  $nM d^{-1}$  (Fig. 2b), with the highest rates at the mid-transect stations 4 and 5, suggesting active  $N_2$  fixation may potentially be restricted by dFe in the east, and P in the west over this transect. This contrasts with a previous west-to-east transect at 10°N which measured highest rates to the east of this transect (16–25°W), corresponding to higher Fe concentrations nearer the coast [83].

In addition to dominating the *nifH* amplicon sequence data, *Trichodesmium* spp. *nifH* genes were detectable by qPCR across the transect, ranging from  $10^4$  to  $10^6$  copies  $L^{-1}$  (Fig. 2c) with a peak ( $3.8 \times 10^6$  copies  $L^{-1}$ ) at station 5 in the middle of the transect, corresponding with one of the highest  $N_2$  fixation rates (11.5  $nM d^{-1}$ ) (Fig. 2c). *Trichodesmium nifH* abundances were also high at stations 4 and 7 ( $2.6 \times 10^5$  copies  $L^{-1}$  and  $2.3 \times 10^6$ , respectively), corresponding with high rates of  $N_2$  fixation.

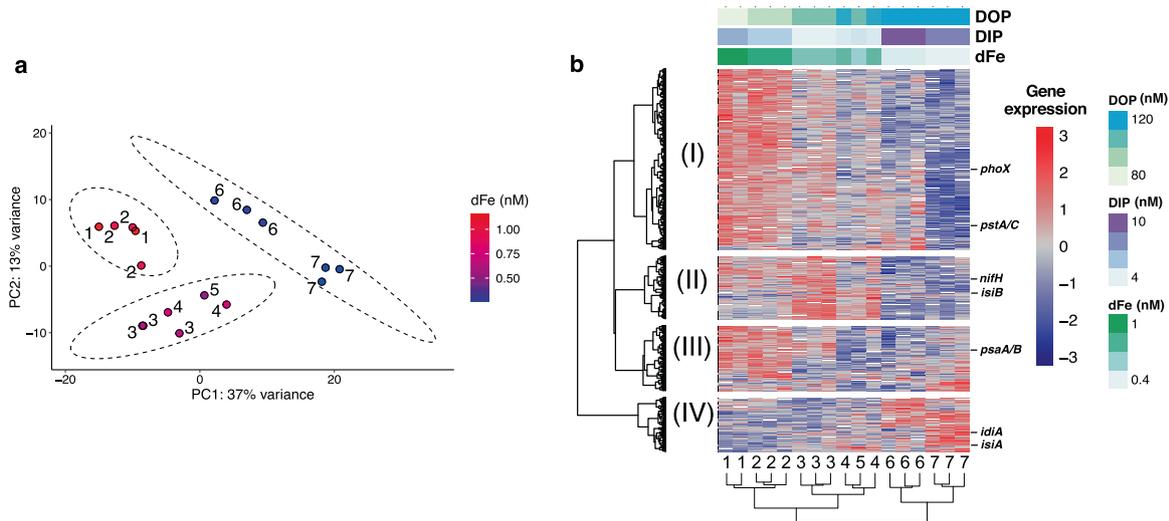
### Functional characterisation of *Trichodesmium*

Metatranscriptomic analysis was performed on hand-picked *Trichodesmium* colonies from stations 1–7 in triplicate. Sequencing was successful on 17 of the 21 samples collected, with an average of 1,208,214 mRNA reads (ranging from 471,088 to 2,260,415 reads) per sample (Supplementary Table 1). OGs of *Trichodesmium* transcripts were examined using a Principal Component Analysis (PCA) on hierarchical associations from Euclidean distances based on expression variability among samples ( $n = 17$ ). Stations clustered into three distinct groups (Fig. 3a), which presented strong reproducibility among biological replicates from each station and coincided with the geographic distribution of stations as well as their characteristic nutrient environments (Fig. 1,

Supplementary Fig. 3). As such, three distinct metatranscriptome regions were defined: west: stations 1–2 where dFe was highest, DOP lowest and hence dFe:TDP was higher; mid-transect: stations 3–5 from the middle of transect where  $N_2$  fixation rates were highest; and east: stations 6–7 where dFe was lowest, DOP highest and hence dFe:TDP lowest. The PCA analysis of the global transcriptome clearly demonstrates that *Trichodesmium* spp. alter gene expression profiles across the observed natural gradients in Fe and P availability.

The transcript analysis across stations revealed significant differential expression for 26.1% (1057 of 4044) of *Trichodesmium* OGs (adjusted  $p$  value  $< 0.1$ ), hereafter termed DEOGs. Spatial DEOG patterns reflected the geographic clusters: (I) upregulated in the west ( $n = 526$ ); (II) up- ( $n = 183$ ) or (III) downregulated in the centre of the gyre ( $n = 190$ ); and (IV) upregulated in the east ( $n = 158$ ) (Fig. 3b; Supplementary Fig. 4). OGs were most differentially expressed between stations 2 and 7 at opposite ends of the gyre (935 differentially expressed OGs in total, representing 23.1% of the total OGs) (Supplementary Table 3). Remarkably, just 3 OGs (0.07%) were differentially expressed between stations 1 and 2 in the west. While the west cluster was distinct from the cluster in the central gyre (stations 3, 4, and 5), expression was fairly conserved in the centre cluster, with a maximum variation of 31 OGs, or 0.77% of the total expressed OGs between stations. The east cluster (stations 6 and 7) showed the greatest within-group diversity, explained by 358 DEOGs (8.8% of the total OGs) between the two stations, of which 61% (219 OGs) had unknown function (not shown). This diversity may reflect species-specific differences in gene expression as *Trichodesmium spiralis* became the dominant *nifH* sequence at station 7 (Fig. 2a), while *Trichodesmium thiebautii nifH* was dominant at all other stations.

To confirm that our conclusions were not affected by our choice of metagenomic reference sequence, this analysis was repeated using the binned *Trichodesmium* metagenomic contigs described in previous work [23, 43]. Frischkorn via personal communication). Mapping rates to this reference were similar, but slightly lower than to the Sargasso Sea reference (37% reads assigned to features vs 42%). Following OG clustering and quantification, a similar transcriptomic profile across the transect was obtained (Supplementary Fig. 5).

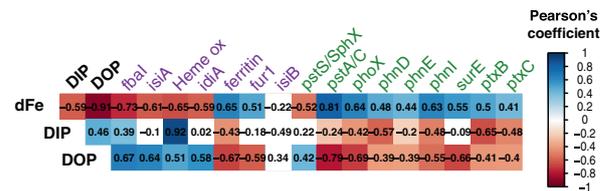


**Fig. 3** *Trichodesmium* metatranscriptome. **a** Principal component analysis (PCA) of all the 4044 *Trichodesmium* OGs across stations from all 17 samples calculated from the variance stabilising transformation (VST) using DESeq2 [68]. Colour gradient reflects measured surface dFe concentrations (0.3–1 nM). Numbers represent the sample stations. **b** Transcript expression patterns of the 1057 most differentially expressed OGs (DEOGs) across stations. Columns are individual samples clustered based on basis of Euclidean distance annotated with station number. Colour scale indicates high (red) to low (blue) DESeq2 VST OG normalised abundance scaled per row. Four clusters of expression patterns (labelled I–IV) were obtained from k-mean optimal clustering by Euclidean distance. Corresponding nutrient concentrations are shown above heatmap (nM).

To further investigate the functions behind these transcriptional differences across regions, DEOGs were clustered into gene functional categories using SEED [56] (Supplementary Fig. 6). Photosynthesis, carbohydrates and N metabolism represented the categories with highest relative OG abundances across stations, in accordance with the main metabolic activities and biogeochemical roles of *Trichodesmium*. From a broad functional perspective, the high dFe:TDP region in the west caused upregulation of P metabolism, cell division and fatty acids. In addition, the high N<sub>2</sub> fixation mid-transect region had upregulation of N metabolism, and downregulation of photosynthesis. Lastly, the low dFe:TDP eastern region had higher membrane transport and Fe metabolism transcription (Supplementary Fig. 6a). However, these functional category levels included a variety of OGs with cosmopolitan distribution throughout the dataset (Supplementary Fig. 6b). Specific photosynthetic OGs identified were differentially regulated throughout transect while subcategories of N metabolism such as N<sub>2</sub> fixation were separate from ammonia or nitrite assimilation (Supplementary Fig. 6b).

### Nutrient stress biomarkers

OGs known to respond to Fe or P availability (Supplementary Table 4) were further analysed to determine the in situ response of *Trichodesmium* populations to the varying nutrient availability across the transect. With the notable exception of flavodoxin (*isiB*) all Fe stress biomarkers measured were significantly correlated (Pearson test,  $p$  value < 0.05) to observed dFe concentrations as expected (Fig. 4). A negative correlation was seen for Fe-stress biomarkers including *idiA/futA*, chlorophyll-binding *isiA* [24, 84, 85], and heme oxygenase, responsible for heme degradation [25]. Further, upregulation of proteins lacking Fe cofactors was seen under the lowest Fe conditions for the non-metal binding fructose-bisphosphate aldolase class-I that may substitute the divalent metal binding fructose-bisphosphate aldolase class-II [21], but was not seen for *isiB* which can replace Fe-dependent ferredoxin in many cyanobacteria [26]. In addition, the Fe-storage protein ferritin [86] and ferric uptake regulators *fur1/2* [87] were positively correlated with dFe, often upregulated under Fe-replete conditions. The availability of Fe thus impacted *Trichodesmium* physiology across the transect. However,



**Fig. 4** *Trichodesmium* nutrient stress. Correlation matrix of normalised counts from selected gene OG associated with Fe (blue) and P (red) metabolism: fructose-bisphosphate aldolase FBA class-I, *isiA*, Heme oxygenase, *idiA*, ferritin, *fur*, *isiB*, *pstS/sphX*, *pstA/C*, *phoX*, *phnD*, *phnE*, *phnI*, *surE*, *ptxB*, *ptxC*; against measured dFe, DOP and DIP concentrations (nM). Pearson correlation coefficient  $r$  value indicated inside boxes ( $n = 17$ ). Colour scale indicates significant positive correlation (violet), negative correlation (green) or non-significant (white) (Pearson parametric correlation test,  $p$  value < 0.05).

transcription of Fe-stress related OGs even at the highest dFe concentrations suggests a response to Fe scarcity is common throughout the system relative to their needs, despite periodic dFe inputs in the NASG [19, 85].

Similar to Fe, all detectable OGs related to P stress, with the exception of one, were significantly negatively correlated with TDP (Pearson test,  $p$  value < 0.05) (Fig. 4). OGs for the high affinity phosphate-specific transporter *pst* were detected across samples. The inner membrane *pstA/C* OG was negatively correlated with DIP, whereas the phosphate-binding *pstS/sphX* OG (usually upregulated under low phosphate conditions), showed no statistical relation to the small observed gradient in DIP (Fig. 4) suggesting it is unresponsive to DIP [33] or that the low DIP concentrations across this transect did not vary enough to induce a significant transcriptional change. Transcripts for OGs involved in phosphite acquisition ( $\text{PO}_3^{3-}$ ) *ptxB* and *ptxC* [36] were also observed to be negatively correlated to DIP. The DIP pool is typically dominated by phosphate rather than phosphite, and therefore we cannot determine any direct responses to phosphite concentration from these data.



from water splitting to a variety of oxidases can represent an Fe-efficient mechanism for generating ATP from PSII activity without any net O<sub>2</sub> production [95]. Interestingly, OGs encoding for three subunits of the cytochrome c oxidases and alternative respiratory oxidases (cox1-3) were upregulated at the mid-transect stations (Fig. 5a), potentially indicating such a mechanism. Similarly, nickel superoxide dismutase OG was significantly upregulated mid-transect (Fig. 5a), potentially indicating a method for scavenging oxygen radicals generated from PSII or alternative oxidases [96]. We thus suggest that under conditions of high cellular N<sub>2</sub> fixation in a Fe-P co-stressed environment, *Trichodesmium* may employ a coordinated metabolic response relying on alternative electron flow strategies involving PSII water-water cycles and elevated *isiA*:PSI ratios to produce ATP [95] while allowing metalloenzyme reallocation strategies yielding Fe atoms from PSI.

Finally, several OGs related to cobalamin were present in the same gene expression profile II and correlated with nitrogenase *nifH* transcripts (Fig. 5a). This further supports the important role of *Trichodesmium* in the mid-transect marine microbiome, including *Trichodesmium* associated bacteria, as a source of both fixed N and essential vitamins [18, 43].

### Adaptation to a constructed Fe-P co-stressed niche

The observed co-regulated expression patterns can be related to the well-established resource ratio theory [9, 14, 97]. The three broad environments along the longitudinal transect defined by Fe and P availability (Supplementary Fig. 3) were associated with the four statistically significant transcriptional profiles (Fig. 3b, Supplementary Fig. 4). Mapping the transcriptional patterns into 'resource-space' based on residual dFe and DOP availability enables definition of these four patterns of gene regulation along this transect. Expression up/downregulation was based on standard deviations variability of z-score scaled expression from each pattern. These are characterised by (I) upregulation when DOP < 90 nM; (II) upregulation when dFe > 0.44 nM and DOP > 90.4 nM; (III) downregulation when dFe > 0.44 nM and DOP > 90.4 nM; and (IV) upregulation when dFe < 0.44 nM (Fig. 6). With transcription profile (II) (Fig. 3b) associated with enhanced N<sub>2</sub> fixation, which was characteristic of the mid-transect stations experiencing Fe and P co-stress [9, 19].

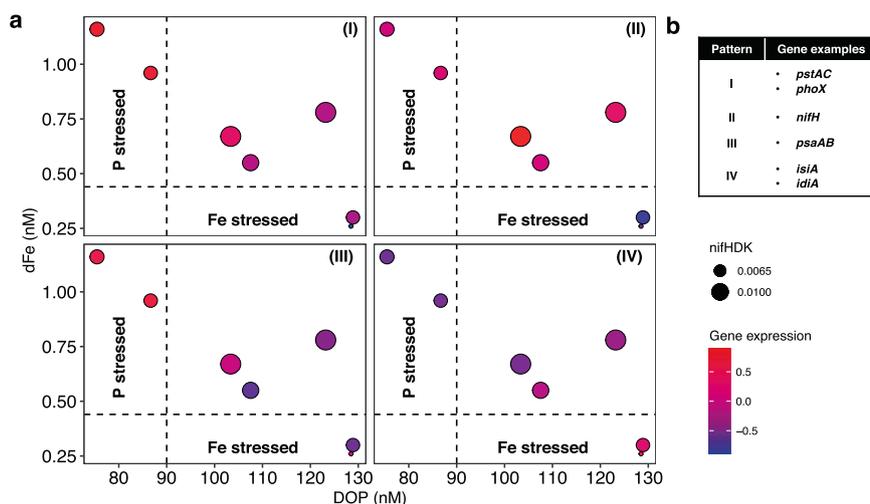
External forcing, including upwelling of waters with excess P (relative to N) and dust-borne Fe inputs, provides the drivers for biogeochemical gradients of the type sampled [13]. However, the

activity of the organisms is also crucial in setting the state of the system, through depleting supplied nutrients to the point where they become limiting [11, 14]. Simple conceptual and numerical models would predict that abrupt shifts between single nutrient limitation patterns for both diazotrophs and non-diazotrophs should occur [9, 14, 97]. In contrast, the data presented here support previous experimental and physiological evidence of more widespread co-stress/limitation [19, 98]. Indeed, expression of both P and Fe stress biomarkers was observed and correlated with Fe and P availability across the whole transect (Figs. 3, 4). Although, as might be expected [14, 97], at the extremes of our transect, transcriptomic profiles indicated reduced nitrogenase expression associated with the highest Fe (east) or P (west) stress (Fig. 6a, b). The observations thus appear to suggest characterisation of the system in terms of large-scale gradients in the severity of co-stress.

Multiple eco-physiological mechanisms have been proposed for development of widespread regions of nutrient co-limitation [88, 99, 100]. Within the studied system we note that our observed transcriptomic patterns (Fig. 3b) suggest considerable metabolic flexibility linked to use of multiple chemical species of co-limiting nutrients, alongside compensation mechanisms which can alter the cellular requirements for these resources, both of which may contribute to enabling broad co-stressed conditions to persist [89]. Thus, in oligotrophic environments where *Trichodesmium* dominates the diazotrophic community, it may be responsible for the development and maintenance of a constructed Fe-P co-stressed niche (Fig. 6). Correspondingly, evolution might be expected to drive *Trichodesmium* towards having the highest growth rates [16, 17] and cellular nitrogenase activity [19] ultimately leading to highest fitness and ecological success under such conditions. However, the apparent lower cellular N<sub>2</sub> fixation potential associated with increased nutrient stresses at either end of the transect suggest that any response to altered external forcing through changes in either P or Fe supply would still likely be a shift in the geographical locations of these niches [11], while enhanced input of either limiting resource would be expected to increase N<sub>2</sub> fixation in a whole system sense.

### Summary

The natural inverse gradients in Fe and P availability captured in this study enabled a holistic view of the in situ molecular response of *Trichodesmium*. We report transcriptomic profiles (based upon



**Fig. 6** Transcription patterns related to resource availability in situ. **a** Transcription patterns (Fig. 3a) mapped into resource space. Dashed lines are  $\pm 2$  standard deviations above the average nutrient concentration of dFe at stations 6–7 and DOP at stations 1–2. Colour scale indicates averaged normalised z-score across stations from the OGs for each of clusters I–IV (see Fig. 3). Symbol sizes indicate the averaged expression for the structural genes of nitrogenase *nifHDK*. **b** Gene examples found on each expression pattern plotted on (a).

multiple genes) suggesting that *Trichodesmium* responds to decreasing Fe or P availability by reducing N<sub>2</sub> fixation capacity. *Trichodesmium* can maintain enhanced N<sub>2</sub> fixation with a consistent and distinct metabolic profile at intermediate Fe-P concentrations, exhibiting a co-stressed expression profile coinciding with enhanced nitrogenase transcription over a large proportion of the North Atlantic gyre. Moreover, coordinated metabolic responses including Fe-reallocation, O<sub>2</sub> reduction and a proposed alternative photosynthetic strategy to generate ATP without the requirement for PSI appear to be associated with enhanced nitrogenase expression in this region. Our results support previous suggestions that *Trichodesmium* has evolved to exploit conditions of Fe-P co-limitation [16, 17, 19], contributing to the global success of this keystone microbe. Co-limitation should thus be considered in ecological modelling of N<sub>2</sub> fixation and understanding of system responses to altered external forcing.

## DATA AVAILABILITY

Data for this study has been deposited in NCBI and can be accessed as BioProject ID PRJNA721670 (metatranscriptome samples) and BioProject ID PRJNA721834 (*Trichodesmium* metagenome). The nutrient concentrations dataset has been submitted to the GEOTRACES Intermediate Data Product (IDP) 2021.

## REFERENCES

- Falkowski PG. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO<sub>2</sub> in the ocean. *Nature*. 1997;387:272–5.
- Zehr JP. Nitrogen fixation by marine cyanobacteria. *Trends Microbiol*. 2011;19:162–73.
- Capone DG, Zehr JP, Paerl HW, Bergman B, Carpenter EJ. *Trichodesmium* a globally significant marine cyanobacterium. *Science* (80-). 1997;276:1221–9.
- Bergman B, Sandh G, Lin S, Larsson J, Carpenter EJ. *Trichodesmium* - a widespread marine cyanobacterium with unusual nitrogen fixation properties. *FEMS Microbiol Rev*. 2013;37:286–302. 37(3):286–302
- Capone DG, Burns JA, Montoya JP, Subramaniam A, Mahaffey C, Gunderson T, et al. Nitrogen fixation by *Trichodesmium* spp.: An important source of new nitrogen to the tropical and subtropical North Atlantic Ocean. *Glob Biogeochem Cycles*. 2005;19:1–17.
- Mahaffey C, Michaels AF, Capone DG. The conundrum of marine N<sub>2</sub> fixation. *Am J Sci*. 2005;305:546–95.
- Moore C, Mills MM, Achterberg EP, Geider RJ, Laroche J, Lucas MI, et al. Large-scale distribution of Atlantic nitrogen fixation controlled by iron availability. *Nat Geosci*. 2009;2:867–71.
- Dyhrman ST, Webb EA, Anderson DM, Moffett JW, Waterbury JB. Cell-specific detection of phosphorus stress in *Trichodesmium* from the Western North Atlantic. *Limnol Oceanogr*. 2002;47:1832–6.
- Snow JT, Schlosser C, Woodward EMS, Mills MM, Achterberg EP, Mahaffey C, et al. Environmental controls on the biogeography of diazotrophy and *Trichodesmium* in the Atlantic Ocean. *Glob Biogeochem Cycles*. 2015;29:865–84.
- Jickells TD, An ZS, Andersen KK, Baker AR, Bergametti C, Brooks N, et al. Global iron connections between desert dust, ocean biogeochemistry, and climate. *Science*. 2005;308:67–71.
- Schlosser CA, Strzpek K, Gao X, Fant C, Blanc É, Paltsev S, et al. The future of global water stress: an integrated assessment. *Earth's Future*. 2014;2:341–61.
- Wu J, Sunda W, Boyle EA, Karl DM. Phosphate depletion in the Western North Atlantic. *Ocean Sci*. 2000;289:759–62.
- Mather RL, Reynolds SE, Wolff GA, Williams RG, Torres-Valdes S, Woodward EMS, et al. Phosphorus cycling in the North and South Atlantic Ocean subtropical gyres. *Nat Geosci*. 2008;1:439–43.
- Ward BA, Dutkiewicz S, Moore CM, Follows MJ. Iron, phosphorus, and nitrogen supply ratios define the biogeography of nitrogen fixation. *Limnol Oceanogr*. 2013;58:2059–75.
- Mills MM, Moore CM, Langlois R, Milne A, Achterberg E, Nachtigall K, et al. Nitrogen and phosphorus co-limitation of bacterial productivity and growth in the oligotrophic subtropical North Atlantic. *Limnol Oceanogr*. 2008;53:824–34.
- García NS, Fu F, Sedwick PN, Hutchins DA. Iron deficiency increases growth and nitrogen-fixation rates of phosphorus-deficient marine cyanobacteria. *ISME J*. 2015;9:238–45.
- Walworth NG, Fu FX, Webb EA, Saito MA, Moran D, McIlvin MR, et al. Mechanisms of increased *Trichodesmium* fitness under iron and phosphorus co-limitation in the present and future ocean. *Nat Commun*. 2016;7:1–11.
- Walworth NG, Fu FX, Lee MD, Cai X, Saito MA, Webb EA, et al. Nutrient-colimited *Trichodesmium* as a nitrogen source or sink in a future ocean. *Appl Environ Microbiol*. 2018;84:1–14.
- Held NA, Webb EA, McIlvin MM, Hutchins DA, Cohen NR, Moran DM, et al. Co-occurrence of Fe and P stress in natural populations of the marine diazotroph *Trichodesmium*. *Biogeosciences* 2020;17:2537–51.
- Polyviou D, Baylay AJ, Hitchcock A, Robidart J, Moore CM, Bibby TS. Desert dust as a source of iron to the globally important diazotroph *Trichodesmium*. *Front Microbiol*. 2018;8:2683.
- Snow JT, Polyviou D, Skipp P, Christmas NA, Hitchcock A, Geider R, et al. Quantifying Integrated Proteomic Responses to Iron Stress in the Globally Important Marine Diazotroph *Trichodesmium*. *PLOS ONE* 2015;10:e0142626.
- Frischkorn KR, Haley ST, Dyhrman ST. Transcriptional and proteomic choreography under phosphorus deficiency and re-supply in the N<sub>2</sub> fixing cyanobacterium *Trichodesmium erythraeum*. *Front Microbiol*. 2019;10:330. 2012;6:1728–39
- Rouco M, Frischkorn KR, Haley ST, Alexander H, Dyhrman ST. Transcriptional patterns identify resource controls on the diazotroph *Trichodesmium* in the Atlantic and Pacific oceans. *ISME J*. 2018;12:1486–95.
- Shi T, Sun Y, Falkowski PG. Effects of iron limitation on the expression of metabolic genes in the marine cyanobacterium *Trichodesmium erythraeum* IMS101. *Environ Microbiol*. 2007;9:2945–56.
- Saito MA, Bertrand EM, Dutkiewicz S, Bulgin VV, Moran DM, Monteiro FM, et al. Iron conservation by reduction of metalloenzyme inventories in the marine diazotroph *Crocospaera watsonii*. *Proc Natl Acad Sci USA* 2011;108:2184–9.
- La Roche J, Boyd PW, McKay RML, Geider RJ. Flavodoxin as an in situ marker for iron stress in phytoplankton. *Nature*. 1996;382:802–5.
- De la Cerda B, Castielli O, Durán RV, Navarro JA, Hervás M, De la Rosa MA. A proteomic approach to iron and copper homeostasis in cyanobacteria. *Brief Funct Genom Proteom*. 2007;6:322–9.
- Chappell PD, Webb EA. A molecular assessment of the iron stress response in the two phylogenetic clades of *Trichodesmium*. *Environ Microbiol*. 2010;12:13–27.
- Polyviou D, Machelett MM, Hitchcock A, Baylay AJ, MacMillan F, Mark Moore C, et al. Structural and functional characterization of *IdiA/FutA* (Tery\_3377), an iron-binding protein from the ocean diazotroph *Trichodesmium erythraeum*. *J Biol Chem*. 2018;293:18099–109.
- Berman-Frank I, Lundgren P, Chen YB, Küpper H, Kolber Z, Bergman B, et al. Segregation of nitrogen fixation and oxygenic photosynthesis in the marine cyanobacterium *Trichodesmium*. *Science*. 2001;294:1534–7.
- Berman-Frank I, Lundgren P, Falkowski P. Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. *Res Microbiol*. 2003;154:157–64.
- Sandh G, Ran L, Xu L, Sundqvist G, Bulone V, Bergman B. Comparative proteomic profiles of the marine cyanobacterium *Trichodesmium erythraeum* IMS101 under different nitrogen regimes. *Proteomics*. 2011;11:406–19.
- Orchard ED, Webb EA, Dyhrman ST. Molecular analysis of the phosphorus starvation response in *Trichodesmium* spp. *Environ Microbiol*. 2009; 11:2400–11.
- Dyhrman ST, Ruttenberg KC. Presence and regulation of alkaline phosphatase activity in eukaryotic phytoplankton from the coastal ocean: Implications for dissolved organic phosphorus remineralization. *Limnol Oceanogr*. 2006;51:1381–90.
- Karl DM. Nutrient dynamics in the deep blue sea. *Trends Microbiol*. 2002;10:410–8.
- Polyviou D, Hitchcock A, Baylay AJ, Moore CM, Bibby TS. Phosphite utilization by the globally important marine diazotroph *Trichodesmium*. *Environ Microbiol Rep*. 2015;7:824–30.
- Obata H, Karatani H, Matsui M, Nakayama E. Fundamental studies for chemical speciation of iron in seawater with an improved analytical method. *Marine Chemistry*. 1997;56:97–106.
- Kunde K, Wyatt NJ, González-Santana D, Tagliabue A, Mahaffey C, Lohan MC. Iron Distribution in the Subtropical North Atlantic: The Pivotal Role of Colloidal Iron. *Glob Biogeochem Cycles*. 2019;33:1532–47.
- Woodward EMS, Rees AP. Nutrient distributions in an anticyclonic eddy in the northeast Atlantic Ocean, with reference to nanomolar ammonium concentrations. *Deep Res Part II Top Stud Oceanogr*. 2001;48:775–93.
- Davis CE, Blackbird S, Wolff G, Woodward M, Mahaffey C. Seasonal organic matter dynamics in a temperate shelf sea. *Prog Oceanogr*. 2019;177:101925.
- Lomas MW, Burke AL, Lomas DA, Bell DW, Shen C, Dyhrman ST, et al. Sargasso Sea phosphorus biogeochemistry: an important role for dissolved organic phosphorus (DOP). *Biogeosci Discuss*. 2009;6:10137–75.
- Klawonn I, Lavik G, Böning P, et al. Simple approach for the preparation of 15–15N<sub>2</sub>-enriched water for nitrogen fixation assessments: evaluation, application and recommendations. *Front Microbiol*. 2015;6:769.

43. Frischkorn KR, Haley ST, Dyhrman ST. Coordinated gene expression between *Trichodesmium* and its microbiome over day-night cycles in the North Pacific Subtropical Gyre. *ISME J*. 2018;12:997–1007.
44. Tang W, Cerdán-García E, Berthelot H, Polyviou D, Wang S, Baylay A, et al. New insights into the distributions of nitrogen fixation and diazotrophs revealed by high-resolution sensing and sampling methods. *ISME J*. 2020;14:2514–26.
45. Zehr JP, McReynolds LA. Use of degenerate oligonucleotides for amplification of the *nifH* gene from the marine cyanobacterium *Trichodesmium thiebautii*. *Appl Environ Microbiol*. 1989;55:2522–6.
46. Zani S, Mellon MT, Collier JL, Zehr JP. Expression of *nifH* genes in natural microbial assemblages in Lake George, New York, detected by reverse transcriptase PCR. *Appl Environ Microbiol*. 2000;66:3119–24.
47. Turk KA, Rees AP, Zehr JP, Pereira N, Swift P, Shelley R, et al. Nitrogen fixation and nitrogenase (*nifH*) expression in tropical waters of the eastern North Atlantic. *ISME J*. 2011;5:1201–12.
48. Hitchen J, Sooknaran R, Khanna A. ScriptSeq V2 Library Preparation Method: A Rapid and Efficient Method for Preparing Directional RNA-Seq Libraries. *J Biomol Tech*. 2012;23:533–534.
49. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet J*. 2011;17:10–2.
50. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. MetaSPAdes: a new versatile metagenomic assembler. *Genome Res*. 2017;27:824–34.
51. Wu Y-W, Tang Y-H, Tringe SG, Simmons BA, Singer SW. MaxBin: an automated binning method to recover individual genomes from metagenomes using. *Microbiome*. 2014;2:4904–9.
52. Parks DH, Imelfort M, Skennerton CT, Hugenholz P, Tyson GW. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*. 2015;25:1043–55.
53. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
54. Hyatt D, Chen GL, LoCascio PF, Land ML, Frank W, Larimer LJH. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinforma*. 2010;11:1–11.
55. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res*. 2002;30:1575–84.
56. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res*. 2014;42:D206–14.
57. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015;12:59–60.
58. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. correspondence QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7:335–6.
59. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26:2460–1.
60. Edgar RC. UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nat Methods*. 2013;10:996–8.
61. Westreich ST, Treiber ML, Mills DA, Korf I, Lemay DG. SAMSA2: a standalone metatranscriptome analysis pipeline. *BMC Bioinforma*. 2018;19:1–11.
62. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
63. Kopylova E, Noé L, Touzet H. SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*. 2012;28:3211–7.
64. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*. 2014;30:614–20.
65. Tatusova T, Ciufu S, Fedorov B, O'Neill K, Tolstoy I. RefSeq microbial genomes database: new representation and annotation strategy. *Nucleic Acids Res*. 2014;42:D553–9.
66. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37:907–15.
67. Liao Y, Smyth GK, Shi W. FeatureCounts: an efficient general-purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30:923–30.
68. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:1–21.
69. Wu S, Mi T, Zhen Y, Yu K, Wang F, Yu Z. A Rise in ROS and EPS Production: New Insights into the *Trichodesmium erythraeum* Response to Ocean Acidification. *J Phycol*. 2021;57:172–82.
70. Sedwick PN, Church TM, Bowie AR, Marsay CM, Ussher SJ, Achilles KM, et al. Iron in the Sargasso Sea (Bermuda Atlantic Time-series Study region) during summer: Eolian imprint, spatiotemporal variability, and ecological implications. *Global Biogeochem Cycles*. 2005;19:GB4006.
71. Hatta MV, Measures CI, Wu J, Roshan S, Fitzsimmons JN, Sedwick P, et al. An overview of dissolved Fe and Mn distributions during the 2010–2011 U.S. GEOTRACES north Atlantic cruises: GEOTRACES GA03. *Deep Res Part II Top Stud Oceanogr*. 2015;116:117–29.
72. Mahaffey C, Reynolds S, Davis CE, Lohan MC. Alkaline phosphatase activity in the subtropical ocean: insights from nutrient, dust and trace metal addition experiments. *Front Mar Sci*. 2014;1:73.
73. Church MJ, Mahaffey C, Letelier RM, Lukas R, Zehr JP, Karl DM. Physical forcing of nitrogen fixation and diazotroph community structure in the North Pacific subtropical gyre. *Global Biogeochem Cycles*. 2009;23:GB2020.
74. Zehr JP, Capone DG. Changing perspectives in marine nitrogen fixation. *Science*. 2020;368:eaay9514.
75. Benavides M, Moisaner PH, Daley MC, Bode A, Aristegui J. Longitudinal variability of diazotroph abundances in the subtropical North Atlantic Ocean. *J Plankton Res*. 2016;38:662–72.
76. Luo YW, Doney SC, Anderson LA, Benavides M, Berman-Frank I, Bode A, et al. Database of diazotrophs in global ocean: Abundance, biomass and nitrogen fixation rates. *Earth Syst Sci Data*. 2012;4:47–73.
77. Moisaner PH, Beinart RA, Voss M, Zehr JP. Diversity and abundance of diazotrophic microorganisms in the South China Sea during intermonsoon. *ISME J*. 2008;2:954–67.
78. Moisaner PH, Serros T, Paerl RW, Beinart RA, Zehr JP. Gammaproteobacterial diazotrophs and *nifH* gene expression in surface waters of the South Pacific Ocean. *ISME J*. 2014;8:1962–73.
79. Robidart JC, Church MJ, Ryan JP, et al. Ecogenomic sensor reveals controls on N<sub>2</sub>-fixing microorganisms in the North Pacific Ocean. *ISME J*. 2014;8:1175–85.
80. Stenegren M, Caputo A, Berg C, Bonnet S, Foster R. Distribution and drivers of symbiotic and free-living diazotrophic cyanobacteria in the western tropical South Pacific. *Biogeosciences*. 2018;15:1559–78.
81. Langlois R, Großkopf T, Mills M, Takeda S, LaRoche J. Widespread Distribution and Expression of Gamma A (UMB), an Uncultured, Diazotrophic,  $\gamma$ -Proteobacterial *nifH* Phylotype. *PLoS ONE*. 2015;10:e0128912.
82. Ratten J-M, LaRoche J, Desai DK, et al. Sources of iron and phosphate affect the distribution of diazotrophs in the North Atlantic. *Deep Sea Res Part II: Topical Stud Oceanogr*. 2015;116:332–41.
83. Voss M, Croot P, Lochte K, Mills M, Peeken I. Patterns of nitrogen fixation along 10°N in the tropical Atlantic. *Geophys Res Lett*. 2004;31:L23509.
84. Bibby TS, Nield J, Barber J. Iron deficiency induces the formation of an antenna ring around trimeric photosystem I in cyanobacteria. *Nature*. 2001;412:743–5.
85. Richier S, Macey AI, Pratt NJ, Honey DJ, Moore CM, Bibby TS. Abundances of iron-binding photosynthetic and nitrogen-fixing proteins of *Trichodesmium* both in culture and in situ from the North Atlantic. *PLoS ONE*. 2012;7:e35571.
86. Keren N, Aurora R, Pakrasi HB. Critical roles of bacterioferritins in iron storage and proliferation of cyanobacteria. *Plant Physiol*. 2004;135:1666–73.
87. González A, Bes MT, Valladares A, Peleato ML, Fillat MF. FurA is the master regulator of iron homeostasis and modulates the expression of tetrapyrrole biosynthesis genes in *Anabaena* sp. PCC 7120. *Environ Microbiol*. 2012;14:3175–87.
88. Sebastian M, Ammerman JW. The alkaline phosphatase PhoX is more widely distributed in marine bacteria than the classical PhoA. *ISME J*. 2009;3:563–72.
89. Browning TJ, Achterberg EP, Yong JC, Rapp I, Utermann C, Engel A, et al. Iron limitation of microbial phosphorus acquisition in the tropical North Atlantic. *Nat Commun*. 2017;8:1–7.
90. Proudfoot M, Kuznetsova E, Brown G, Rao NN, Kitagawa M, Mori H, et al. General enzymatic screens identify three new nucleotidases in *Escherichia coli*: Biochemical characterization of SurE, Yfbr, and YjgG. *J Biol Chem*. 2004;279:54687–94.
91. Orchard ED, Benitez-Nelson CR, Pellechia PJ, Lomas MW, Dyhrman ST. Polyphosphate in *Trichodesmium* from the low-phosphorus Sargasso Sea. *Limnol Oceanogr*. 2010;55:2161–9.
92. Berman-Frank I, Cullen JT, Shaked Y, Sherrell RM, Falkowski PG. Iron availability, cellular iron quotas, and nitrogen fixation in *Trichodesmium*. *Limnol Oceanogr*. 2001;46:1249–60.
93. Schoffman H, Keren N. Function of the IsiA pigment–protein complex in vivo. *Photosynth Res*. 2019;141:343–53.
94. Küpper H, Ferimazova N, Setlik I, Berman-Frank I. Traffic lights in *Trichodesmium*. Regulation of photosynthesis for nitrogen fixation studied by chlorophyll fluorescence kinetic microscopy. *Plant Physiol*. 2004;135:2120–33.
95. Behrenfeld MJ, Milligan AJ. Photophysiological expressions of iron stress in phytoplankton. *Ann Rev Mar Sci*. 2013;5:217–46.
96. Ho TY. Nickel limitation of nitrogen fixation in *Trichodesmium*. *Limnol Oceanogr*. 2013;58:112–20.
97. Tilman D. Resources: a graphical-mechanistic approach to competition and predation. *Am Nat*. 1980;116:362–3.

98. Mills MM, Ridame C, Davey M, La Roche J, Geider RJ. Iron and phosphorus co-limit nitrogen fixation in the eastern tropical North Atlantic. *Nature* 2004;429:292–4.
99. Saito MA, McIlvin MR, Moran DM, Goepfert TJ, DiTullio GR, Post AF, et al. Multiple nutrient stresses at intersecting Pacific Ocean biomes detected by protein biomarkers. *Science* 2014;345:1173–7.
100. NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group. Moderate-resolution Imaging Spectroradiometer (MODIS) Aqua Chlorophyll Data. MODIS-Aqua Level 3 Mapped Chlorophyll Data Version R2018.0. NASA OB.DAAC, Greenbelt, MD, USA. Published online 2017.

## ACKNOWLEDGEMENTS

We would like to thank the Captain and crew of the RRS James Cook JC150, as well as K. Kunde, N. Wyatt, J. Harley, N. Held and S. Sargent for sampling and nutrient analysis assistance. This work was supported by the UK NERC funded project “Zinc Iron and Phosphorus Co-Limitation in the Ocean (ZIPOc) (NE/N001079/1) to CM and ML that provided nutrient analysis, N<sub>2</sub> fixation rates and ancillary oceanographic data. Supporting funding for molecular sampling and analysis came from UK NERC grant MARINeDNA (NE/N006151/1) and EU-H2020 Project AtlantOS (Grant Agreement No. 633211) (JCR). ECG is the recipient of a Graduate School of the National Oceanography Centre Southampton PhD studentship.

## AUTHOR CONTRIBUTIONS

ECG and DP collected samples while at sea. ECG, DP, and JCR processed samples, did qPCR, *nifH* and metatranscriptomic sequencing with AB. AB prepared and sequenced libraries, prepared and sequenced metagenomic samples and developed the bioinformatics approach. ECG and AB analysed the sequence data. CM, ML, MW, LW contributed with nutrients/rate measurements and funding for oceanographic data collection. JCR funded the molecular work. ECG, JCR, TSB, and CMM analysed the data. ECG wrote the paper, with significant contributions from TSB, JCR, and CMM.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41396-021-01151-1>.

**Correspondence** and requests for materials should be addressed to E. Cerdan-Garcia or J. C. Robidart.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021