



# Dynamic diel proteome and daytime nitrogenase activity supports buoyancy in the cyanobacterium *Trichodesmium*

Noelle A. Held<sup>1,2,6</sup>, John B. Waterbury<sup>3</sup>, Eric A. Webb<sup>4</sup>, Riss M. Kellogg<sup>1,2</sup>, Matthew R. McIlvin<sup>1</sup>, Michael Jakuba<sup>5</sup>, Frederica W. Valois<sup>3</sup>, Dawn M. Moran<sup>1</sup>, Kevin M. Sutherland<sup>1,2,7</sup> and Mak A. Saito<sup>1</sup>✉

**Cyanobacteria of the genus *Trichodesmium* provide about 80 Tg of fixed nitrogen to the surface ocean per year and contribute to marine biogeochemistry, including the sequestration of carbon dioxide. *Trichodesmium* fixes nitrogen in the daylight, despite the incompatibility of the nitrogenase enzyme with oxygen produced during photosynthesis. While the mechanisms protecting nitrogenase remain unclear, all proposed strategies require considerable resource investment. Here we identify a crucial benefit of daytime nitrogen fixation in *Trichodesmium* spp. that may counteract these costs. We analysed diel proteomes of cultured and field populations of *Trichodesmium* in comparison with the marine diazotroph *Crocospaera watsonii* WH8501, which fixes nitrogen at night. *Trichodesmium*'s proteome is extraordinarily dynamic and demonstrates simultaneous photosynthesis and nitrogen fixation, resulting in balanced particulate organic carbon and particulate organic nitrogen production. Unlike *Crocospaera*, which produces large quantities of glycogen as an energy store for nitrogenase, proteomic evidence is consistent with the idea that *Trichodesmium* reduces the need to produce glycogen by supplying energy directly to nitrogenase via soluble ferredoxin charged by the photosynthesis protein PsaC. This minimizes ballast associated with glycogen, reducing cell density and decreasing sinking velocity, thus supporting *Trichodesmium*'s niche as a buoyant, high-light-adapted colony forming cyanobacterium. To occupy its niche of simultaneous nitrogen fixation and photosynthesis, *Trichodesmium* appears to be a conspicuous consumer of iron, and has therefore developed unique iron-acquisition strategies, including the use of iron-rich dust. Particle capture by buoyant *Trichodesmium* colonies may increase the residence time and degradation of mineral iron in the euphotic zone. These findings describe how cellular biochemistry defines and reinforces the ecological and biogeochemical function of these keystone marine diazotrophs.**

Biological nitrogen (N<sub>2</sub>) fixation is a crucial biogeochemical process, being foundational to marine microbial food webs and primary productivity in N-limited ecosystems<sup>1–6</sup>. A limited number of marine microorganisms can fix nitrogen, and they have distinct yet overlapping niches in the N-limited surface ocean<sup>7,8</sup>. The ecological drivers differentiating the habitats of diverse marine nitrogen fixers are uncertain but critical for understanding the current and future global N budget<sup>1</sup>. As a result, our ability to resolve environmental and biochemical controls on nitrogen fixation rates remains limited<sup>9–12</sup>.

Here we examine an important yet enigmatic marine nitrogen fixer, *Trichodesmium*, which contributes substantial amounts of fixed nitrogen to the ocean annually, and due to vertical export may disproportionately affect nitrogen and carbon cycling at depth<sup>12–15</sup>. *Trichodesmium* performs nitrogen fixation and photosynthesis simultaneously during the photoperiod, and this is perplexing because the nitrogenase enzyme is susceptible to damage by molecular oxygen produced by photosystem II<sup>16</sup>. It is unclear how *Trichodesmium* protects the sensitive nitrogenase enzyme from damage by molecular oxygen. In general, photosynthesizing

diazotrophs have evolved two main strategies for solving this problem. Some, such as *Crocospaera*, separate the processes temporally, fixing carbon during the day and nitrogen at night<sup>17–19</sup>. Others, such as *Anabaena*, separate the processes spatially, forming differentiated nitrogen-fixing heterocyst cells that lack the oxygen-evolving photosystem II complex<sup>20–23</sup>. While it has been hypothesized that *Trichodesmium* forms partially differentiated ‘diazocyte’ cells for the purpose of nitrogen fixation, the evidence for this strategy remains mixed and controversial<sup>24–31</sup>. Fine-tuned temporal separation of peak nitrogen fixation and photosynthetic activity has also been suggested<sup>26,32</sup>, although it is also possible for the processes to occur simultaneously<sup>33</sup>. Regardless of the exact mechanisms, it is clear that daytime nitrogen fixation requires considerable cellular coordination and resource investment to protect the vulnerable nitrogenase enzyme.

This study therefore addresses a key question – why does *Trichodesmium* fix nitrogen during the day? Since *Trichodesmium* is ecologically successful and widely distributed in the marine environment, the benefits of daytime nitrogen fixation must outweigh the costs involved in protecting the nitrogenase enzyme from

<sup>1</sup>Marine Chemistry and Geochemistry Department, Woods Hole Oceanographic Institution, Woods Hole, MA, USA. <sup>2</sup>Department of Earth, Atmospheric and Planetary Science, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>3</sup>Department of Biology, Woods Hole Oceanographic Institution, Woods Hole, MA, USA. <sup>4</sup>Marine and Environmental Biology, Department of Biological Sciences, University of Southern California, Los Angeles, CA, USA. <sup>5</sup>Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, MA, USA. <sup>6</sup>Present address: Department of Environmental Systems Science, ETH, Zurich, Switzerland. <sup>7</sup>Present address: Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA. ✉e-mail: [msaito@whoi.edu](mailto:msaito@whoi.edu)

degradation by oxygen. Since cyanobacterial physiology, in particular nitrogen fixation and photosynthesis, is strongly dependent on daily rhythms, we investigated the proteomes of the model organism *T. erythraeum* IMS101, as well as field populations from the Atlantic Ocean during the diel cycle to reveal key biochemical insights on nitrogen and carbon fixation. We then compared the diel proteome of *Trichodesmium* to the previously published diel proteome of *Crocospaera watsonii* WH8501, which fixes nitrogen at night<sup>17</sup>. Together, these high-resolution datasets provide an opportunity to probe the molecular physiology of this unique nitrogen fixer, and to understand the intersection of biochemical activity that allows *Trichodesmium* species to occupy their important niche in the marine ecosystem.

### Whole proteome dynamics of *Trichodesmium erythraeum* IMS101 vs *Crocospaera watsonii* WH8501

The diel proteome of *Trichodesmium erythraeum* IMS101 was explored by high-resolution global proteomics analysis of triplicate cultures maintained in a 14:10 light/dark cycle, with ramped dawn and dusk transitions mimicking those in subtropical gyres. Samples were collected every 1–3 h for proteomics, biomass Carbon, Hydrogen, and Nitrogen (CHN) content analysis and nitrogenase activity via the acetylene reduction assay<sup>34</sup>. The cultures were grown in RMP media supplied with ample phosphorus and trace-metal nutrients<sup>35</sup>. Exponential growth (0.7–0.8 doublings per day) was observed on the basis of total protein content (Extended Data Fig. 1b). The *Trichodesmium* filaments appeared healthy, and colonies, which in *T. erythraeum* are indicative of halted growth, nutrient stress and/or redox stress<sup>34,36–39</sup>, were not observed<sup>40</sup>.

The proteomics analysis provided a global look at the molecular physiology of the *Trichodesmium* cultures. In total, 2,492 proteins were identified, representing 56% of the *T. erythraeum* IMS101 predicted protein encoding genes (Supplementary Table 1). Protein abundances are reported as normalized spectral abundance factor (NSAF)-normalized spectral counts to control for protein size and small variation in the total protein (~1 µg) injected across the samples. The NSAF normalization therefore allows relative protein abundances to be compared across the experiment. Total spectral counts did not vary systematically across the experiment, indicating consistency in proteome depth across the analytical runs (Extended Data Fig. 1a). However, based on a colorimetric assay, total protein content per volume of culture was greater at night, indicative of night-time protein accumulation and/or higher protein degradation and turnover during the day, consistent with the enrichment of protease and peptidase enzymes during the photoperiod (Extended Data Fig. 1b and Supplementary Fig. 2).

*Trichodesmium's* proteome was highly dynamic and key systems such as nitrogen fixation, photosynthesis, central carbohydrate metabolism and ribosomal proteins oscillated four times and as much as 200% within a few hours (Fig. 1a,b). Major proteome components included phycobilisome proteins, growth-related proteins (ribosomes and chaperonin GroEL), nitrogenase and the photosystems (Supplementary Table 2 and Fig. 3). Nitrogen fixation, light-driven electron transport, photosynthesis and related processes, including ATP synthesis (dark blue cluster), clustered separately from most other functional modules (green cluster), indicating temporal separation of nitrogen fixation/photosynthesis from other metabolic activities (Fig. 1a). The offsetting of these proteome clusters can be clearly seen in Fig. 1b, where the general patterns of protein abundance in the clusters are mirror images of one another. There was no general increase or decrease in protein abundances during the diel cycle that would indicate culture effects during the 24 h experiment (Extended Data Fig. 2). The surprisingly rapid dynamics of the proteome was reproduced in three independent proteomic observations: the main experiment described above, a separate replicate experiment conducted one year earlier (Extended

Data Fig. 3 and Supplementary Table 3) and in field populations of *Trichodesmium* sampled over a diel cycle on 10 March 2018 from the subtropical Eastern Atlantic (65°22.420' W, 17°0.284' N; Extended Data Fig. 4 and Supplementary Table 4). *Trichodesmium* has a large number of two component regulatory systems, proteases and peptidases, which were abundant, cycled continuously throughout the day and were probably involved in regulating these rapid proteome changes (Supplementary Table 2 and Fig. 2).

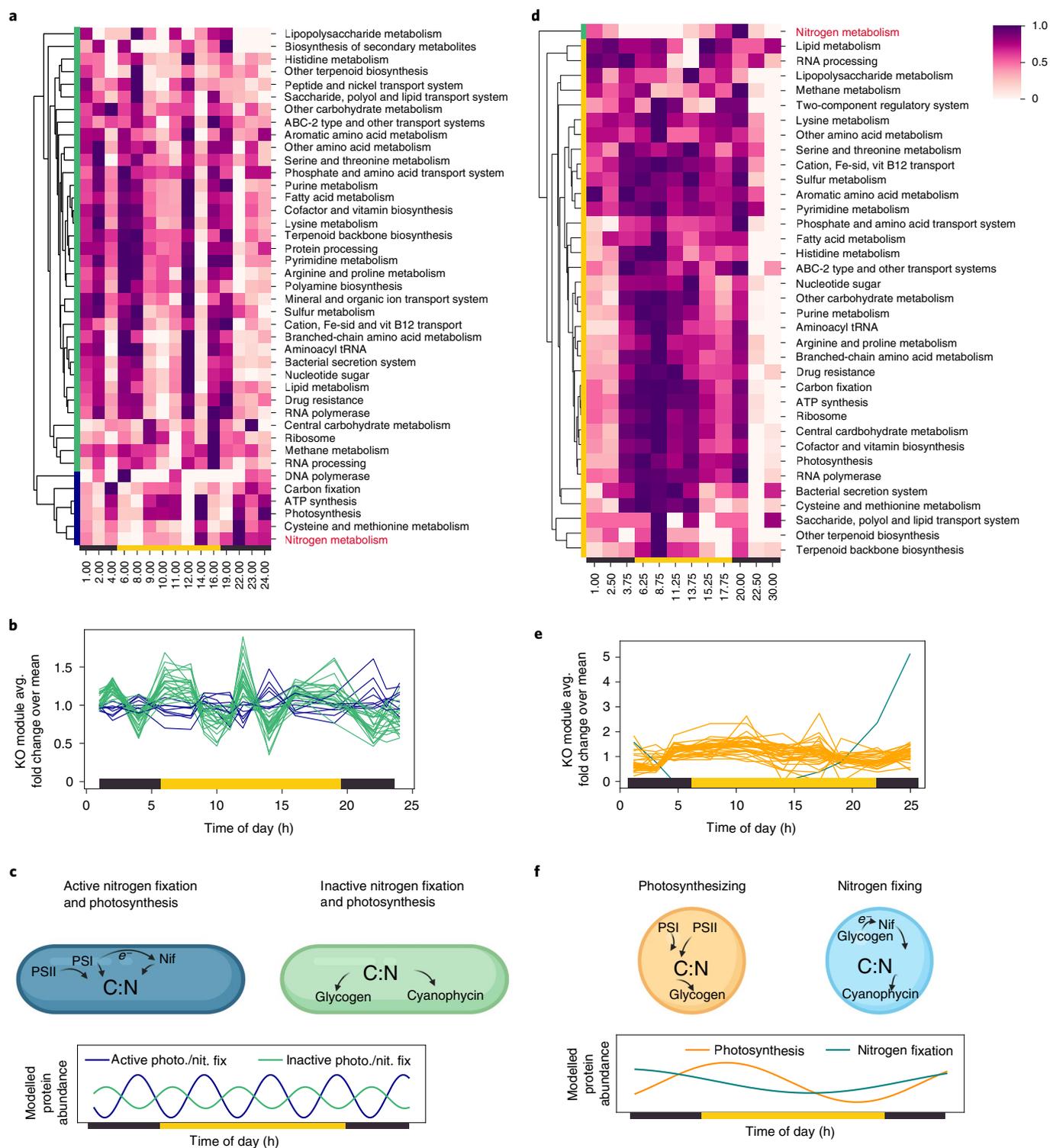
The overall dynamics of *Trichodesmium's* proteome differed from that of *Crocospaera*, highlighting the unique nitrogen fixation strategy of *Trichodesmium*. In *Trichodesmium*, the proteome oscillated four times throughout the diel period, and was as variable at night as during the day (Fig. 1a–c). By contrast, *Crocospaera's* proteome had only one complete oscillation, with most proteins being abundant in the photoperiod, and nitrogenase being most abundant at night, consistent with dark nitrogen fixation (Fig. 1d–f). A single proteome oscillation based on light/dark cycling, and in particular the temporal separation of photosynthesis and nitrogenase proteins, is common in unicellular diazotrophs and is in stark contrast to the rapid oscillation of the *Trichodesmium* proteome<sup>41</sup>. We interpret the variation in *Trichodesmium's* proteome to reflect the need to accommodate nitrogen fixation, photosynthesis and/or other processes during the photoperiod, and note that the timescale of the observed changes suggests a crucial role for rapid regulatory processes, such as protein post-translational modification, which were not profiled in this experiment and could lead to lower signals at the protein level. Yellow and dark purple bars indicate the light and dark periods, respectively.

### Nitrogen fixation and photosynthesis occur simultaneously during the photoperiod in *Trichodesmium*

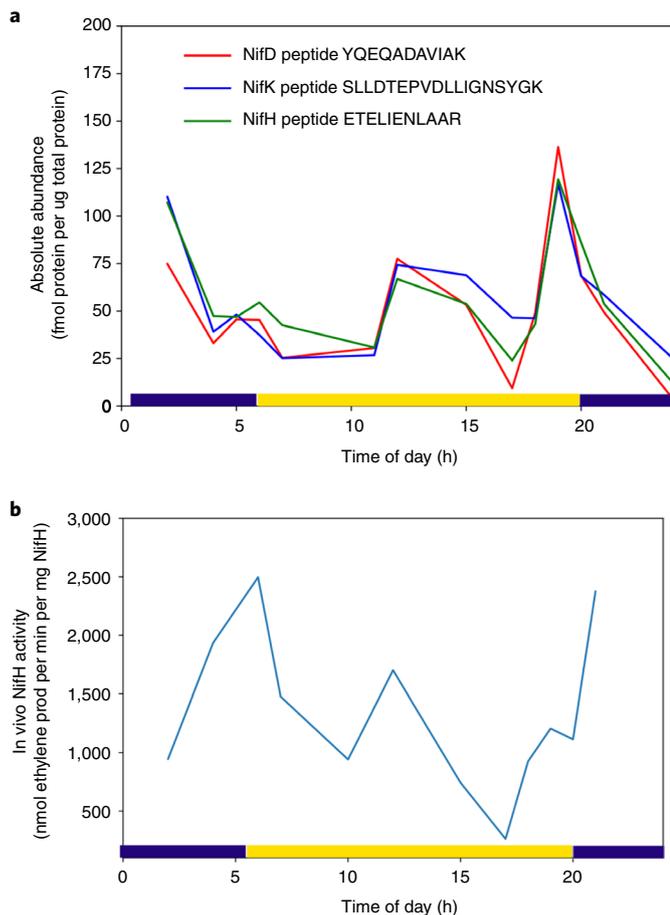
While in *Crocospaera* nitrogen fixation was clearly temporally separated from photosynthesis, in *Trichodesmium* three independent lines of evidence demonstrated that these processes occur simultaneously during the photoperiod: nitrogenase protein concentrations, measured nitrogen fixation rates, and diel changes in particulate organic carbon (POC) and particulate organic nitrogen (PON) content.

First, photosynthesis and nitrogenase enzymes were highly abundant during the photoperiod and appeared coherently in the same protein cluster, suggesting that these processes are co-regulated (dark blue cluster in Fig. 1a,b). The possibility of direct regulatory links between nitrogenase and the photosystem proteins is supported by similar observations of co-regulation in the metaproteomes of field populations sampled across geographic space<sup>42</sup>. Again, the clustering of the nitrogen fixation and photosynthesis proteins was reproducible in diel proteomes of *Trichodesmium* field populations and in a separate laboratory replicate conducted one year before the main experiment (Extended Data Figs. 3 and 4). The diel pattern of nitrogenase protein abundance differed from previously reported patterns in nitrogenase transcripts, which generally increase during the day and decrease at night<sup>30</sup>. In the proteomes, all major subunits of the nitrogenase enzyme oscillated 3–4 times throughout the diel period, and were most abundant in the early morning and mid-afternoon, with a dip at mid-day. Interestingly, we also observed a late-afternoon spike in nitrogenase abundance, which was consistent across the replicates and in the field proteomes. These patterns were confirmed by absolute quantitation of the nitrogenase NifH, NifD and NifK subunits using <sup>15</sup>N-labelled peptide standards (Fig. 2a and Supplementary Table 5). Unlike *Crocospaera*, which completely degrades and re-synthesizes its nitrogenase pool each day, in *Trichodesmium*, nitrogenase never fully disappeared and was abundant even when nitrogen fixation rates inferred from acetylene reduction/ethylene production rates were below detection.

The second line of evidence was provided by acetylene reduction/ethylene production rates, an analytical proxy for nitrogen fixation, which demonstrated that nitrogen fixation occurred mainly in the



**Fig. 1 | Dynamics of the *Trichodesmium* proteome in comparison to *Crocosphaera*.** **a**, Heatmap of the *T. erythraeum* IMS101 proteome during the diel cycle. Proteins were gathered into Kegg Ontology (KO) modules, summed for each time point, averaged across the three biological replicates, then unit-normalized across the row and clustered by UPGMA. Two clusters emerged, indicated by green and dark blue bars on the left hand side. Yellow and black bars at the bottom of the heatmap indicate the light and dark periods, and numbering is hours after midnight. See Supplementary Fig. 1 for a version of this heatmap where the clustering algorithm was also applied to the x axis (time of day). **b**, The same data with the abundance of each KO module presented as a line and coloured on the basis of the clustering in **a**. **c**, Summary of major physiological modes emergent from the diel proteome of *Trichodesmium*, with colours consistent with **a** and **b**, and including active photosynthesis/nitrogen fixation (dark blue) and inactive photosynthesis/nitrogen fixation (green). Modelled protein abundances over the diel cycle were generated by fitting sinusoidal functions to the summed protein abundances in each cluster. Blue model coefficient of determination  $R^2 = -5.7$ , green model  $R^2 = -2.9$ . **d-f**, The same information as in **a-c** but for *Crocosphaera watsonii* WH8501 (data from ref. <sup>34</sup>). Here, the major emergent physiological modes were photosynthesizing (orange) and nitrogen fixing (turquoise). Turquoise model  $R^2 = 0.62$ , orange model  $R^2 = 0.63$ . PS-I, photosystem I; PS-II, photosystem II; Nif, nitrogenase.



**Fig. 2 | Nitrogenase concentration and activity over the diel cycle.**

**a**, Absolute abundance of each of the main nitrogenase enzyme subunit proteins, calibrated using  $^{15}\text{N}$ -labelled standards. **b**, In vivo specific activity rate for NifH over the diel cycle, calculated by dividing acetylene reduction/ethylene production rates (see Fig. 3c) by the measured concentration of the nitrogenase enzyme. Variation in the specific activity implies post-translational control of enzymatic activity<sup>30,44</sup>. Yellow and dark purple bars indicate light and dark periods, respectively.

photoperiod, particularly in the early morning just after dawn. This approach was combined with the absolute protein concentration measurements to calculate nitrogenase-specific activity rates throughout the diel period. The observed in vivo specific activity varied between 261 and 1,934 nmol ethylene per mg NifH per min, and was greatest in the early morning hours, particularly after dawn (Fig. 2b). The specific activity is similar to those reported for other nitrogen fixers (900–1,200 nmol ethylene per mg NifH per min for *Azotobacter vinelandii*, scaling with growth rate)<sup>43</sup>. The variation in nitrogenase activity contrasted with that of *Crocospaera*, which exhibited a clearer on/off signal, with nitrogenase being present and highly active only at night (Extended Data Fig. 5). Together, these observations are consistent with previous evidence that the nitrogenase iron protein NifH in *Trichodesmium* is regulated by post-translational modification, and/or that energy availability regulates nitrogenase activity<sup>30,44</sup>. The ability to maintain an intact pool of nitrogenase and modulate its activity by post-translational modification or other means may explain how diel patterns of nitrogen fixation are observed to respond to environmental conditions such as light availability<sup>45</sup>,  $p\text{CO}_2$ <sup>46,47</sup>, temperature<sup>48</sup> and oxygen concentrations<sup>33</sup>.

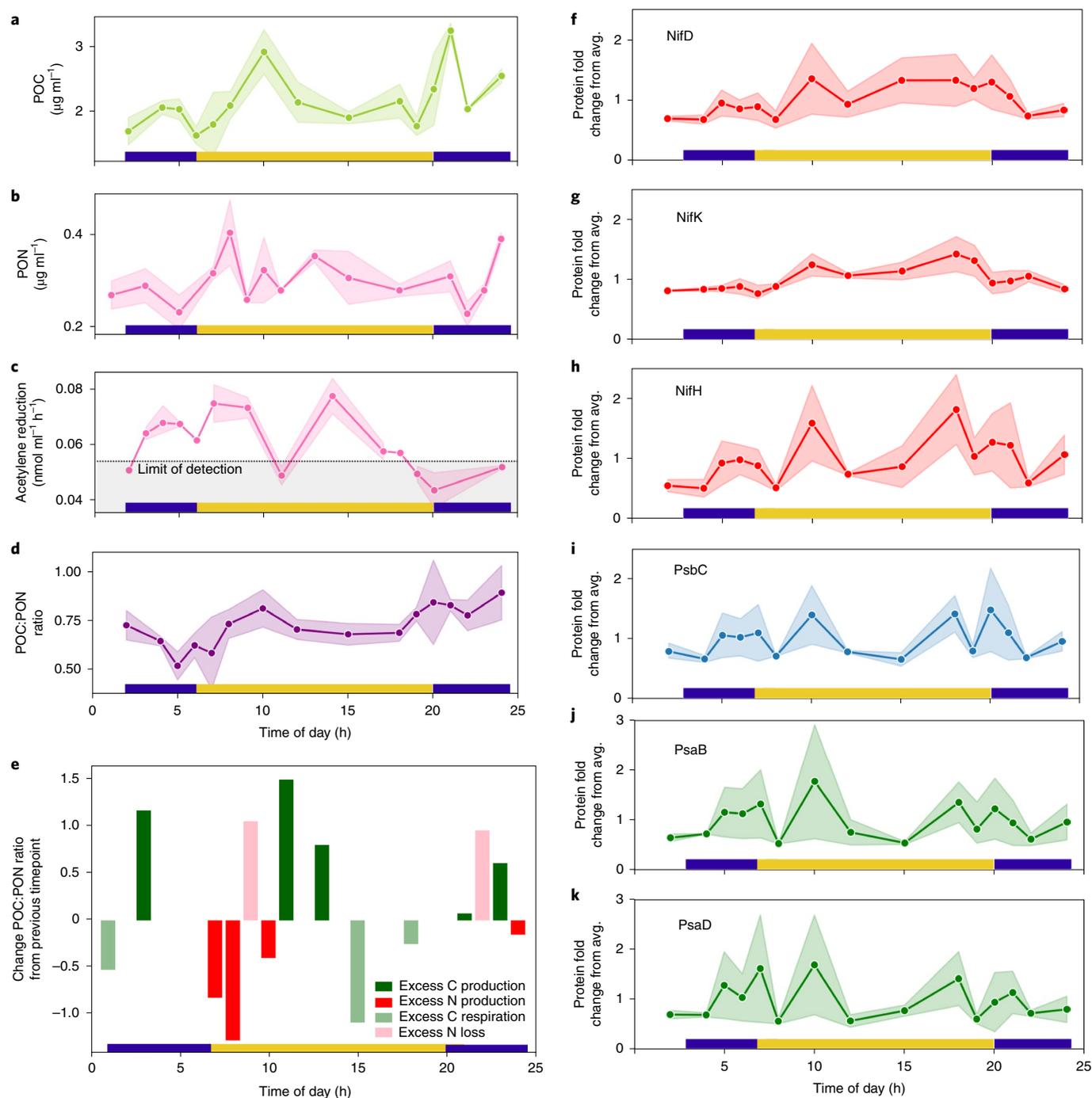
The third line of evidence for simultaneous photosynthesis and nitrogen fixation was provided by temporal changes in POC

(Fig. 3a) and PON (Fig. 3b), which were balanced throughout the day (Supplementary Table 6). As neither fixed nitrogen nor carbon were provided to the cultures, changes in POC content can be attributed to carbon fixation and respiration, and changes in PON content attributed to nitrogen fixation and loss. The relative balance of POC versus PON content can be used to infer excess production: excess carbon fixation occurs when POC and POC:PON increase, excess nitrogen fixation occurs when PON increases and POC:PON decreases, excess carbon respiration occurs when POC and POC:PON decrease, and excess nitrogen loss occurs when PON decreases and POC:PON increases. We also considered situations in which multiple processes occurred, that is, decreased POC:PON ratio associated with carbon respiration and nitrogen production, or increased POC:PON ratio associated with carbon production and nitrogen loss, but this was not observed. Based on this framework, excess nitrogen fixation occurred in the early morning, followed by excess carbon fixation around noon, and then consistency in the POC:PON ratio for much of the afternoon (Fig. 3c–e; see also Supplementary Figs. 4 and 5 and Table 7, which reports the results of an additional replicate consistent with Fig. 3). This indicates simultaneous and generally balanced bulk production of POC and PON through coupled photosynthesis and nitrogen fixation – possibly a key benefit of daytime nitrogenase activity. POC content was only weakly correlated with protein content ( $r^2=0.13$ , Extended Data Fig. 6), consistent with previous reports that diel POC content is driven by changes in carbohydrate abundance, and that total protein content is relatively constant<sup>49</sup>. Relative stability in the POC:PON ratio of *Trichodesmium* has been observed previously<sup>48,50</sup> and implies that carbon fixation rates can be used to estimate nitrogen production by *Trichodesmium*<sup>51</sup>.

Together, the nitrogenase protein concentrations, acetylene reduction rates and POC/PON content clearly demonstrated daytime nitrogen fixation activity. However, whereas in some studies nitrogen fixation peaks at mid-day concurrent with decreased photosynthetic efficiency<sup>14,24,26</sup>, here mid-day was associated with a transient decrease in acetylene reduction rates (Fig. 3c), consistent with the observed decrease in nitrogenase and photosystem protein abundance (Fig. 2a and Fig. 3f–k), nitrogenase specific activity (Fig. 2b) and cellular PON content (Fig. 3b). The lack of consensus regarding the diel nitrogen fixation patterns of *Trichodesmium* is not surprising given multiple observations of its dependence on environmental conditions. Indeed, diel patterns of nitrogen fixation based on acetylene reduction measurements vary in the literature, and the observed peak occurs at different times ranging from mid-morning to mid-afternoon, indicating that diel nitrogen fixation and its regulation is not well-understood<sup>24,33,45,52–54</sup>.

### Distinct roles for glycogen and nitrogenase in *Trichodesmium* vs *Crocospaera*

Glycogen production is an integral part of the ecological strategy of Cyanobacteria, and particularly serves as a means to regulate stress due to nitrogen limitation or high-light exposure. Glycogen serves two major and related purposes, first as an energy storage compound and second to prevent damage to the photosystems by providing a safe sink for excess electrons when carbon fixation is hindered, for instance by lack of nitrogen<sup>55,56</sup>. *Crocospaera* benefit from these functions by producing large amounts of glycogen during the day and consuming the stored energy to fuel nitrogenase activity at night; therefore, glycogen content in *Crocospaera* is positively correlated with photosynthesis protein abundance and negatively correlated with nitrogenase abundance<sup>17</sup>. In contrast, glycogen synthesis protein GlcG in *Trichodesmium* was strongly and negatively correlated with both the photosynthesis and nitrogen fixation proteins (Spearman correlation coefficient =  $-0.91$ ,  $P=1 \times 10^{-16}$ ) (Fig. 4a–e and Supplementary Table 8), suggesting a fundamentally different role for glycogen production in the

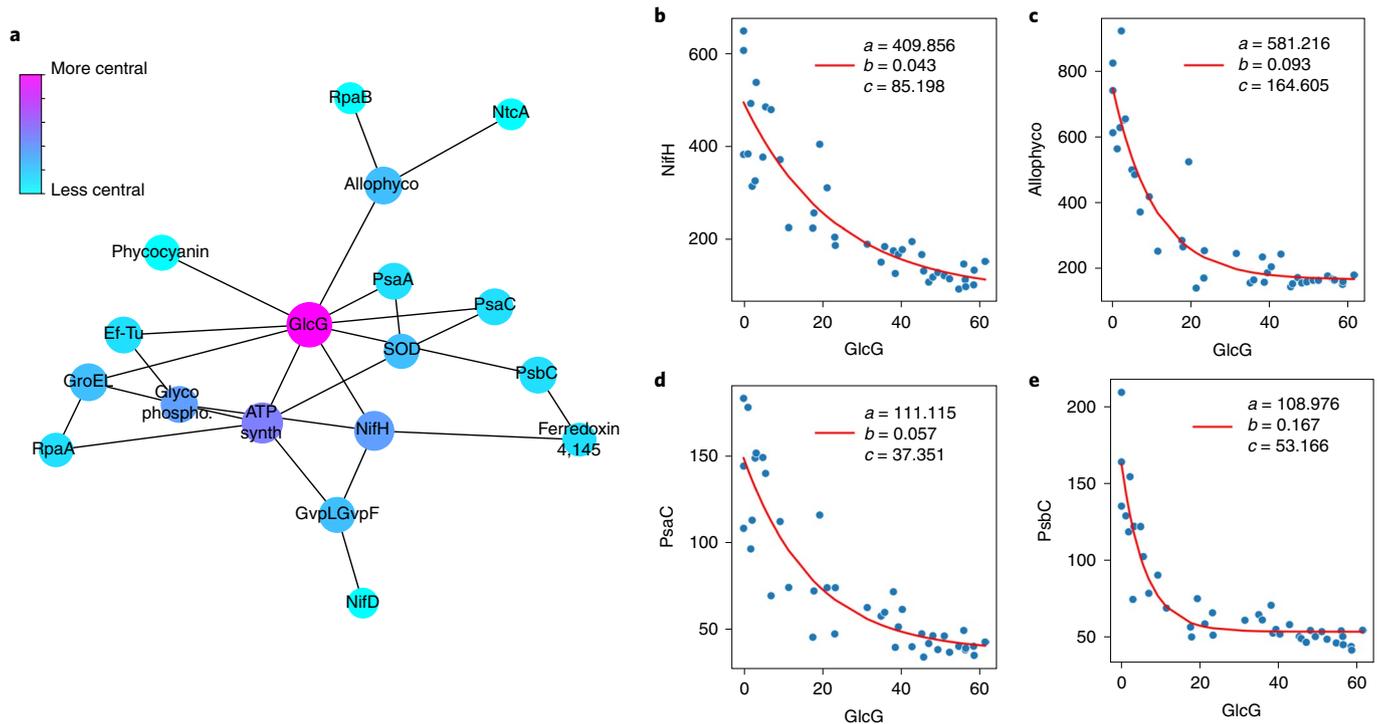


**Fig. 3 | Temporal dynamics of biomass composition, nitrogen fixation rates, and associated proteins in *Trichodesmium*.** **a–d**, Temporal dynamics of POC (**a**), PON (**b**), acetylene reduction rates (**c**) and the POC:PON ratio (**d**) in cultured *T. erythraeum* over the diel cycle. **e**, Temporal changes in the POC:PON ratio relative to previous timepoint, with bars colour coded on the basis of associated changes in either POC or PON content, with the possible processes being excess C production, excess N production, excess C respiration or excess N loss (see text). **f–k**, Associated temporal changes in the abundance of the three nitrogenase subunits (**f–h**), photosystem I protein PsbC (**i**) and photosystem II proteins PsaB and PsaD (**j,k**). Each dot represents the average value across the biological triplicates, and shaded areas represent 90% confidence intervals for the replicates, calculated by bootstrapping ( $n=1,000$ ). Protein abundances are normalized to their average value across the entire diel cycle. Yellow and dark purple bars indicate the light and dark periods, respectively.

physiology of *Trichodesmium* that is related to its daytime nitrogen fixation patterns.

We suggest that the positive correlation between nitrogenase and the photosystems, together with the negative correlation of these enzymes to the glycogen synthesis protein GlcG, indicates that *Trichodesmium* uses the energy-demanding nitrogenase enzyme,

which consumes 16 ATPs and 8 electrons per reaction turnover<sup>57</sup>, in lieu of glycogen to modulate energy during the photoperiod. In this way, bursts of photosynthesis at defined intervals during the photoperiod may be linked to bursts of nitrogen fixation that not only relieve N-limitation, ensuring balanced POC and PON production, but also prevent damage to the photosystems by consuming excess



**Fig. 4 | A glycogen synthesis protein emerges as a hub in the *Trichodesmium proteome*.** **a**, Network of negatively correlated photosynthesis and nitrogen fixation related proteins. Edges are drawn for protein pairs with Spearman rank-order correlation coefficients  $< -0.8$  and correlation  $P$  values  $< 0.05$ . Node size and colour designate relative degree of closeness centrality in the network. **b-e**, Examples of negative protein abundance correlations used to build the network, shown here as a function of the central protein in the network, GlcG. Each point represents one independent biological observation, that is, one of the replicates at one of the diel sampling timepoints. Curves were fit to the general exponential function  $ae^{(-bx)} + c$  (red lines) via least-squares regression, and constants for the best fit are provided in the legend.

light energy. The photosynthesis protein PsaC was positively correlated with the photosystem enzymes and nitrogenase, and may be the conduit for direct energy transfer between photosystem II and nitrogenase via soluble ferredoxin (Fig. 4d). In this way, *Trichodesmium* may be distinct from *Crocospaera*, which always uses the non-iron containing flavodoxin to provide electrons to nitrogenase to reduce cellular iron demand<sup>17</sup>. If electrons are supplied to nitrogenase via a soluble ferredoxin charged by PsaC, this would add to the realized iron requirement of the nitrogenase enzyme. However, the mechanism of electron transport to nitrogenase may be flexible depending on iron conditions, since *Trichodesmium* is known to substitute flavodoxin for ferredoxin during Fe stress, and here only high iron conditions were considered<sup>58,59</sup>.

### Daytime nitrogen fixation allows *Trichodesmium* to occupy a high-light niche

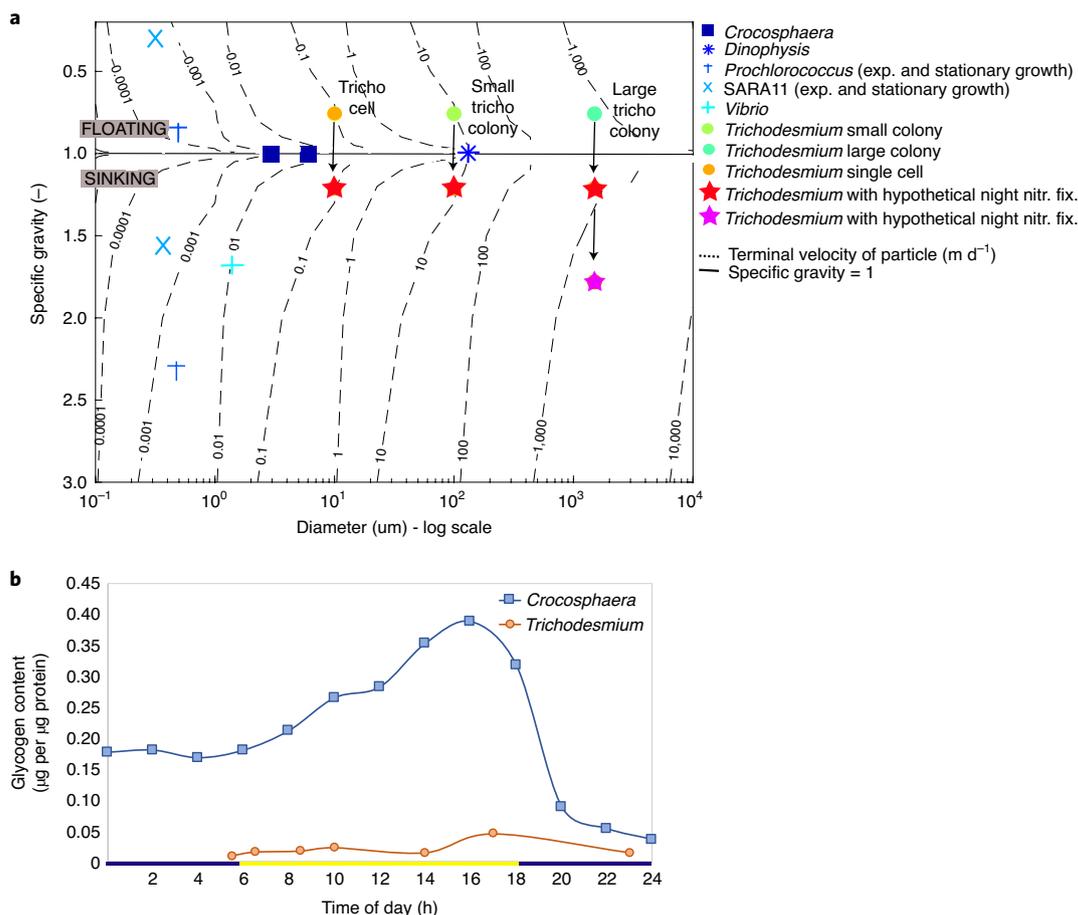
Daytime nitrogen fixation allows the direct shunt of electrons from the photosystems to nitrogenase, and may reduce glycogen production and the density of the cell, thereby helping *Trichodesmium* to remain neutrally buoyant. *Trichodesmium* cell density is clearly impacted by the amount of carbohydrate ballast in the cell, and glycogen represents a large fraction of this ballast<sup>49,60–62</sup>. To assess the benefits of daytime nitrogen fixation to *Trichodesmium*, we introduced a hypothetical night nitrogen fixation scenario in which, similar to *Crocospaera*, the majority of nitrogen fixation is conducted in the dark (Supplementary Table 9). In this hypothetical night nitrogen fixation scenario, *Trichodesmium* would have to produce glycogen to provide energy to nitrogenase, which would increase the POC content of the cell by ~60% (Table 1 and Supplemental Box 1). This would most probably cause *Trichodesmium* cells to sink, given the high density of glycogen ( $0.5\text{--}2\text{ g ml}^{-1}$ )<sup>63</sup> and the

**Table 1 | Calculated benefits and costs of normal daytime nitrogen fixation to *Trichodesmium*, compared with the hypothetical night nitrogen fixation scenario described in the text**

Maximum POC content	
Actual day nitrogen fixation scenario ( $\mu\text{g ml}^{-1}$ culture)	3.976
Hypothetical night nitrogen fixation scenario ( $\mu\text{g ml}^{-1}$ culture)	6.362
% increase hypothetical night nitrogen fixation vs actual day nitrogen fixation	60%
Minimum Fe bound to inactive nitrogenase	
Actual day nitrogen fixation scenario ( $\mu\text{mol mol}^{-1}$ POC)	10
Hypothetical night nitrogen fixation scenario ( $\mu\text{mol mol}^{-1}$ POC)	$\sim 0^*$

Details of the calculations are provided in the supplementary material. POC content is higher in the hypothetical night nitrogen fixation scenario to fuel the energy-demanding nitrogenase enzyme in the absence of light energy. Cellular Fe requirements may also differ – night-time nitrogen fixation allows for ‘hot-bunking’ between the nitrogenase and photosynthesis enzymes since the processes are separated temporally (\*).

observation that *Trichodesmium* already invests heavily in buoyancy maintenance by producing gas vesicles, the proteins for which are highly expressed (Supplementary Table 2), and which occupy up to 70% of the cell volume<sup>60</sup>. Indeed, the glycogen content of *Trichodesmium* is much lower than in *Crocospaera*, indicating reduction of ballast in *Trichodesmium* cells (Fig. 5b). Furthermore, the glycogen content was higher in *Trichodesmium* cells sampled at

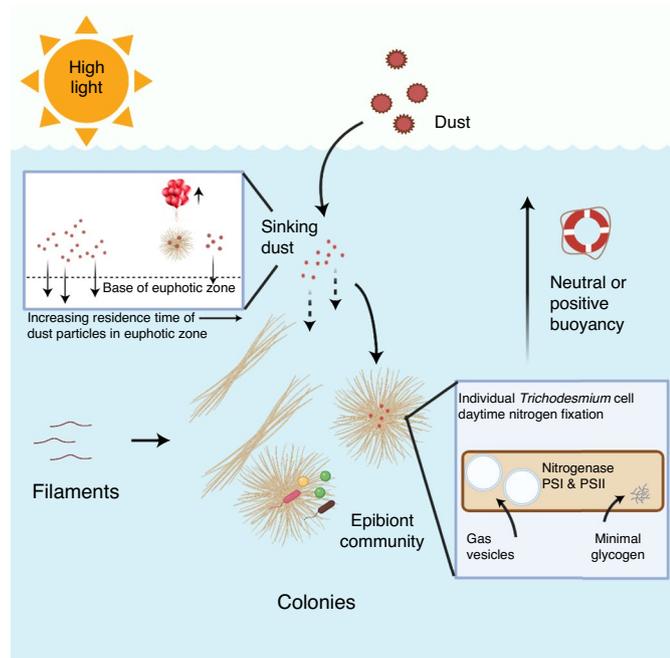


**Fig. 5 | Buoyancy properties of *Trichodesmium* and other phytoplankton. a**, The terminal velocities of phytoplankton particles of various characteristic diameters and specific gravities relative to seawater of density  $1,025 \text{ kg m}^{-3}$  (see Supplementary Table 9). Black contours indicate the value of the particle's terminal velocity in  $\text{m d}^{-1}$ , with negative values indicating upwards, floating velocity (that is, when the specific gravity is less than 1), and positive values indicating downwards, sinking velocity (that is, when the specific gravity is greater than 1). For a given species, cell density can differ depending on growth conditions (for example, *Prochlorococcus* and SARA11, each displayed in exponential (sinking) and stationary (floating) growth). The density of the *Trichodesmium* particles was modelled using literature values and allowing 25% of the cell volume to be occupied by a gas vesicle, resulting in positive buoyancy. The hypothetical dark nitrogen fixation case increases the density of the *Trichodesmium* biomass by 60%, resulting in sinking. Arrows are used to indicate the change in terminal velocity for normal light nitrogen-fixing and hypothetical dark nitrogen-fixing *Trichodesmium* particles. For large colonies carrying dust particles, the additional mass of iron ( $1.03 \mu\text{g Fe per } 1\text{-mm-diameter colony}$ ) further exacerbates the sinking velocity. **b**, Glycogen content per  $\mu\text{g}$  total protein for *Trichodesmium* and *Crocospaera* cells over the diel cycle. (*Crocospaera* glycogen data from ref. <sup>17</sup>). Yellow and dark purple bars indicate the light and dark periods, respectively.

depth, demonstrating that increased glycogen content is associated with sinking (Extended Data Fig. 7).

In contrast to small spheroid *Crocospaera* cells, *Trichodesmium*'s efforts to maintain neutral buoyancy are compounded by its large cell size and tendency to aggregate in filamentous colonies. This is related to the dependence of the drag force on the Reynolds number, a non-dimensional parameter dependent on speed and shape that exerts a dominant influence on the drag experienced by an object moving through a fluid. We calculated the terminal velocity of spherical marine microbial particles of characteristic diameters and specific gravities selected from the literature, including individual *Trichodesmium* cells, and small and large puff-type *Trichodesmium* colonies (see Supplementary Table 9). The sinking speed was predicted from a force balance of buoyancy and drag, where the drag coefficient was allowed to vary as a function of Reynolds number, itself dependent on speed, and the solution found iteratively. The sign of the terminal velocity, dependent on the calculated specific gravity of the particle relative to seawater, indicated whether the particle was expected to sink (downwards velocity) or to float (upwards

velocity). *Trichodesmium* particles were modelled such that 25% of the cell volume was occupied by a low-density gas vesicle, resulting in initial floating velocities for single cells and colonies (Fig. 5a). This is consistent with previous observations that a substantial fraction of *Trichodesmium* colonies are positively or at least neutrally buoyant, depending on their carbohydrate content<sup>62</sup>. The hypothetical dark nitrogen fixation scenario, described above, was modelled by increasing the biomass component of the *Trichodesmium* specific gravity estimation by 60%, as described in Table 1. In all cases, the added glycogen density resulted in a switch from floating to sinking for the *Trichodesmium* particles, indicating that use of nitrogenase at night could fundamentally change the buoyant properties of the organism. For large *Trichodesmium* colonies, the estimated sinking speed for the hypothetical dark nitrogen fixation case exceeded  $100 \text{ m d}^{-1}$ , meaning that the particles would leave the euphotic zone very rapidly. It should be noted that for tractability, we modelled the particles as smooth spheres; surface roughness, particularly 'spiky' morphology in puff-type colonies, could increase drag and slow the sinking speed<sup>60</sup>. These results thus additionally highlight the



**Fig. 6 | Summary of physiological properties and behaviors that are reinforced by daytime nitrogen fixation.** *Trichodesmium* cells, filaments and colonies have a tendency to sink due to their large size and tendency to aggregate. Daytime nitrogen fixation minimizes glycogen ballast. *Trichodesmium* can also regulate buoyancy through production of gas vesicles that confer neutral and/or positive buoyancy. Thus, *Trichodesmium* is more likely to remain in the upper water column, where they benefit from their high-light-adapted pigments, their ability to access iron from dust particles and interactions with the epibiont community. Additionally, buoyant *Trichodesmium* colonies may increase the residence time of dust particles in the euphotic zone, providing more time for particle solubilization. Not drawn to scale.

potential importance of colony morphology on the biogeochemical function of *Trichodesmium*, in particular their impact on carbon and nitrogen export to depth<sup>12</sup>. Supporting this, we observed that in field populations, the glycogen content of 'spiky' *Trichodesmium* puff colonies was greater than that of *Trichodesmium* tuft colonies (Extended Data Fig. 8).

Thus, whereas glycogen provides an essential means to store energy for night-time nitrogen fixation in unicellular diazotrophs, glycogen ballast presents a problem in *Trichodesmium*. Shifting nitrogen fixation to the day minimizes glycogen production and provides a means for *Trichodesmium* colonies to stay afloat. In this way, *Trichodesmium*'s large size and tendency to form colonies may have driven it to daytime nitrogen fixation or vice versa – by helping *Trichodesmium* to remain at the surface, the development of daytime nitrogen fixation patterns may have allowed *Trichodesmium* cells to develop social behaviours, including the formation of large colonies. It is unclear whether these behaviours evolved simultaneously or sequentially, and if so, which came first. However, it is clear that daytime nitrogen fixation is crucial to *Trichodesmium*'s niche as a high-light, neutrally buoyant, large-celled colonial diazotroph.

### The iron cost of daytime nitrogen fixation

While clearly beneficial for buoyancy, daytime nitrogen fixation also comes at a cost, in particular driving the high iron quota of *Trichodesmium* cells. As mentioned before, *Crocospaera* is able to reduce its daily iron demand by about 40% by completely degrading and re-synthesizing nitrogenase, allowing it to 'hot-bunk' iron atoms

and repurpose them for photosynthesis during the day<sup>17</sup>. Assuming that nitrogenase does not exist as an apoenzyme, the rapid cycling of nitrogenase and the fact that it is never fully degraded mean that *Trichodesmium* cannot benefit from this iron-sparing strategy as *Crocospaera* does. In the quantitative proteomic analysis of the NifH enzyme, the lowest observed concentration of nitrogenase was 27 fmol  $\mu\text{g}^{-1}$  total protein (midnight) and this increased 5.5-fold to a maximum of 116 fmol  $\mu\text{g}^{-1}$  at mid-morning (Fig. 2 and Supplementary Table 5). Assuming normal nitrogenase subunit configuration (one NifH dimer + one NifD/NifK heterotetramer) and full loading of the iron-binding sites (15 atoms per enzyme), this is equivalent to a minimum of 187 fmol iron per  $\mu\text{g}$  total protein at midnight and a maximum of 877 fmol iron per  $\mu\text{g}$  total protein at the height of nitrogen fixation. Using measured ratios of total protein:biomass POC content and the molecular weight of glucose as a conservative proxy for POC, this is equivalent to 10–39  $\mu\text{mol}$  Fe bound to nitrogenase per mol biomass carbon, representing at least 30% of the cellular iron quota, consistent with a previous estimation (Table 1)<sup>64,65</sup>. Thus, the need to maintain simultaneously active pools of nitrogenase and photosynthetic machinery probably drives the high iron requirement of *Trichodesmium* cells, consistent with the frequent observations of iron stress in natural populations<sup>42,58,66</sup>.

To occupy its niche of simultaneous nitrogen fixation and photosynthesis, *Trichodesmium* appears to be a conspicuous consumer of iron, and has developed specific iron-acquisition strategies, some of which depend on colony morphology, to do so. One particularly relevant iron-acquisition behaviour is the ability to acquire iron from particulate sources. Particulate-iron utilization represents a benefit to colony formation and occurs via specialized mechanisms, including biologically mediated reduction of particulate Fe(III) to Fe(II) and possible involvement of siderophores produced by the epibiont community<sup>67–70</sup>. Daytime nitrogen fixation is a key driver of *Trichodesmium*'s high iron quota, but it also provides a means for colonies to maintain buoyancy, and may therefore enhance access to atmospheric dust particles that enter at the surface. Given that staying afloat is a problem for *Trichodesmium* colonies, it is interesting that they have adapted to carry dense mineral particles, that is, they are literally carrying small rocks. From previously published synchrotron-based element maps of a *Trichodesmium* puff, we estimated that a ~1-mm-diameter colony carried 1.03  $\mu\text{g}$  of iron in the form of mineral oxides and silicates<sup>70</sup>. Assuming that the density of the *Trichodesmium* biomass was similar to that of literature observations, this results in a 56% increase in the density of the colony and a significant increase in the sinking velocity, on the order of 1,000  $\text{m d}^{-1}$  in the hypothetical dark nitrogen fixation case (Fig. 5 and Extended Data Fig. 9). Thus, the development of particle-utilization behaviours may reinforce the need for ballast reduction via daytime nitrogen fixation patterns in *Trichodesmium*, as well as the production of gas vesicles to keep the colony afloat.

By effectively slowing the sinking speed of mineral particles, *Trichodesmium* colonies may enhance the retention and residence time of particulate minerals in the euphotic zone (Fig. 6 left inset). This would provide more time for solubilization of iron from these minerals, hence gas-vesicle-containing, positively buoyant *Trichodesmium* colonies may play important roles in the biogeochemical cycling of particulate metals. In this way, daytime nitrogen fixation and dust utilization reinforce the specific niche of *Trichodesmium* and underlie its abundance in high-dust surface ocean ecosystems<sup>3,69,71,72</sup>. Figure 6 summarizes how these multiple aspects of *Trichodesmium*'s cellular biochemistry together define its particular ecological niche.

### Implications for the diazocyte hypothesis

*Trichodesmium* is often pointed to as a unique case among nitrogen fixers because it performs nitrogen fixation and photosynthesis simultaneously without the use of differentiated heterocyst cells<sup>73,74</sup>.

Spatial separation of nitrogen fixation into partially differentiated 'diazocytes' was proposed in *Trichodesmium*<sup>26,75</sup>; however, the evidence for this strategy has been inconsistent and unrepeatable<sup>27,29,31</sup>. Despite the lack of further experimental evidence for diazocytes, there has been confusion in the literature resulting in the assumption of spatial separation as a major cost associated with daytime nitrogen fixation<sup>57,76</sup> (see Supplemental text for a brief review of the current evidence). Our study suggests that a key benefit of daytime nitrogen fixation is the ability to channel light-derived energy directly to the nitrogenase enzyme, thus reducing glycogen ballast and helping *Trichodesmium* to remain buoyant. It is unclear how this finding could be reconciled with the diazocyte theory, since this would require large amounts of energy to be transferred from the vegetative to the diazocyte cells. In true heterocyst-forming Cyanobacteria, reduced carbon is passed to the heterocysts as sucrose, but the *Trichodesmium* genome lacks the crucial sucrose metabolism proteins.

The molecular mechanisms of the hypothesized diazocyte formation are similarly unclear; the involvement of the regulatory protein HetR has been suggested, as this protein is crucial to heterocyst formation in the true heterocyst-forming cyanobacterium *Anabaena* and the protein is present in *Trichodesmium*. However, the hetR gene is widely distributed among filamentous Cyanobacteria (including non-nitrogen-fixing Cyanobacteria), evolved before heterocysts and is probably involved in other processes, including filament cooperation more broadly<sup>77</sup>. Furthermore, we note that the *T. erythraeum* genome lacks other necessary heterocyst formation and maintenance genes<sup>78</sup>, such as hetN and patS. Thus, although our results do not necessarily preclude the existence of diazocytes, we find no evidence for them and instead suggest that the dynamic cycling of the proteome throughout the diel cycle could support nitrogen fixation without the need to invoke spatial separation.

## Conclusions

Daytime nitrogen fixation by *Trichodesmium* requires considerable investment in the regulation, synthesis and degradation of photosystem components, nitrogenase and other metabolic proteins on short timescales. Here we identify the niche-defining benefits that seem to counteract these costs. In summary, the ability to directly shuttle energy from the photosystems to the nitrogenase enzyme reduces glycogen ballast, allowing *Trichodesmium* filaments and colonies to remain at the surface. There, *Trichodesmium* is uniquely poised to access iron from atmospherically derived dust particles, driving its competitiveness in high-dust environments and underscoring the potential importance of buoyant *Trichodesmium* colonies in the processing of mineral particles in the euphotic zone<sup>9,79</sup>. These findings build on previous frameworks relating nutrient acquisition to cell size and shape, where it is generally thought that smaller cells are better competitors in the oligotrophic ocean, owing to their larger surface area:volume ratios<sup>42,80,81</sup>. Specifically, this study highlights that being small is not the only way that a cyanobacterium can maximize its viability in oligotrophic environments, but rather that a large-celled colonial cyanobacterium can become highly competitive in an oligotrophic system by adapting novel biochemical patterns that reinforce each other to define a specialized niche. Importantly, it is not just a single aspect of *Trichodesmium's* biochemistry and physiology that provides these benefits, but rather a collection of connected cellular processes, which together defines the ecology and biogeochemical impact of these unique nitrogen fixers.

## Methods

**Culture conditions.** *Trichodesmium erythraeum* IMS101 was grown in RMP growth medium prepared with oligotrophic Sargasso seawater<sup>35</sup>. The cultures were not axenic but had been sterile-transferred for hundreds of generations; few epibiont proteins were identified, suggesting the cultures were strongly dominated

by *T. erythraeum*. The cultures were maintained for months at 26.9 °C in a 14:10 day:night light cycle with light ramp up/down at dusk and dawn mimicking conditions at Station ALOHA. Approximately 200 ml of dense, healthy, replicate cultures were inoculated into triplicate experimental vessels in a final volume of 1,200 ml RMP media with constant, gentle stirring and gentle oxygenation. The cultures were allowed to acclimate and grow for 5 d before subsampling by sterile pipetting every 1–3 h. Distinct samples ( $n=45$ ) were collected onto 0.2 µm filters for proteomics (50 ml), GFF filters for CHN analysis (10 ml) and for acetylene reduction assay (20 ml, 1 h incubation period). A separate replicate experiment was conducted in the same way one year earlier (see Supplementary Fig. 4 and Extended Data Fig. 3). The doubling time calculated from changes in total protein concentration was 0.66–0.76 per day, and protein content represented approximately 20% of the total biomass POC. CHN analyses were performed by combustion on an elemental analyser (Control Equipment, model CEC 440HA) at the UCSB Marine Science Institute Analytical Laboratory.

**Acetylene reduction assay.** The acetylene reduction assay was conducted by injecting 2 ml concentrated acetylene gas into the headspace of a sealed 50 ml Nalgene culture bottle containing 20 ml of the sampled culture. The culture was returned to the incubator for 1 h, then the reduction product (ethylene) was measured by gas chromatography on a Shimadzu GC-8A and calibrated to a 9.1 ppb ethylene standard.

**Protein digestion and metaproteomic analysis.** Proteins were analysed by a global proteomics/DDA method. In global proteomics, proteins are broken into small pieces, analysed and then reassembled bioinformatically. The protein targets are not selected ahead of time. Proteins were extracted by an SDS detergent-based method and trypsin-digested in a polyacrylamide tube gel<sup>42,82,83</sup>. Protein abundance, which was used to estimate doubling time, was quantified by a colorimetric BCA protein assay (Thermo Fisher). Peptide mixtures were concentrated to 1 µg protein µl<sup>-1</sup> solution and 10 µl or 10 µg was injected per analysis. The global proteomes were analysed by Reverse Phase Liquid Chromatography - active modulation - Reverse Phase Liquid Chromatography Mass Spectrometry (RPLC-am-RPLC-MS)<sup>84</sup> involving two orthogonal in-line chromatography steps (PLRP-S (200 µm × 150 mm, 3 µm bead size, 300 Å pore size, NanoLCMS Solutions), followed by C18 column (100 m × 150 mm, 3 µm particle size, 120 Å pore size, C18 Repronil-God, Maisch, packed in a New Objective PicoFrit column)) on a Thermo Dionex Ultimate 3000 LC system equipped with two RSLC pumps. The first dimension of chromatography was an 8 h pH 10 gradient (10 mM ammonium formate and 10 mM ammonium formate in 90% acetonitrile), which was trapped on alternating dual traps and eluted at 30 min intervals at 500 nl min<sup>-1</sup> onto the C18 column (0.1% formic acid and 0.1% formic acid in 99.9% acetonitrile). The resulting eluent was injected directly onto a Thermo Orbitrap Fusion mass spectrometer with a Thermo Flex ion source. MS1 scans were monitored between  $m/z$  380 and 1,580, with an  $m/z$  1.6 MS2 isolation window (CID mode), 50 ms maximum injection time and 5 s dynamic exclusion time.

Raw spectra were searched using the SequestHT algorithm implemented in Proteome Discoverer 2.2. Peptide sequences were mapped to a *Trichodesmium erythraeum* IMS101 genome (RefSeq NC\_008312.1) plus non-redundant non-*Trichodesmium* sequences identified in a recent metatranscriptome analysis of epibiont organisms associated with *T. erythraeum* IMS101 cultures<sup>85</sup>. Search parameters were 10 ppm parent mass tolerance, 0.8 Da fragment mass tolerance, cysteine modification +57.022 and methionine modification +16. Protein identifications were made via PeptideProphet implemented in Scaffold (Proteome Software) with a stringent 0.01% peptide and 0.3% protein global false discovery rates. Epibiont protein identifications were very sparse and inconsistent, so only *Trichodesmium* proteins were considered in the downstream analysis. Relative quantitation was done by normalized spectral count using the NSAF method implemented in Scaffold. The NSAF normalization controls for sample-to-sample variation in amount of material injected into the mass spectrometer, as well as variation in protein sequence length<sup>86</sup>. Normalized spectral counts can therefore be used to compare protein abundances across samples in the dataset.

**Absolute quantitation of the nitrogenase enzyme.** Absolute quantitation of the nitrogenase enzyme was conducted by parallel reaction monitoring<sup>87</sup> using an isotopically labelled standard for each peptide quantified. The heavy-labelled standard peptide was generated by expressing it in a custom-designed plasmid in competent tuner(DE3)pLys *Escherichia coli* cells growing in <sup>15</sup>N-labelled medium<sup>42</sup>. The over-expressed standard was isolated in inclusion bodies, trypsin-digested and calibrated using a known amount of commercially available Pierce standard peptide mixture (catalogue number 88320), the peptides of which were also included in the custom plasmid. The standard deviation of this calibration was approximately 10%. Linearity of the calibration within the range of expected experimental peptide concentrations was confirmed by a dilution curve based on precursor ion intensity and covering concentrations in the range 0.001 to 200 fmol µl<sup>-1</sup> (Extended Data Fig. 10). Experimental samples were prepared at 0.1 µg µl<sup>-1</sup> total experimental protein and contained the heavy-labelled peptide standard at a concentration of 10 fmol µl<sup>-1</sup>; 10 µl of the sample was injected into the mass spectrometer. The liquid chromatography settings were similar to the

metaproteomic analysis, but only a single dimension of chromatography (C18 column) and a 2 h chromatography gradient was used. The mass spectrometer was run in parallel reaction monitoring mode such that targeted precursor ions ( $m/z$ ) were isolated for fragmentation and MS<sup>2</sup> analysis. Peptide quantitation was performed in Skyline<sup>88</sup> using the top 6 most abundant fragment ions for each peptide. The ratios of the heavy (standard) vs light (experimental) MS<sup>2</sup> peak areas were averaged and this was used to calculate the nitrogenase concentration, accounting for protein extraction efficiency, with corrections for measured protein extraction efficiency (Supplementary Table 6). Absolute nitrogenase concentrations are reported as femtomoles per microgram of total protein.

**Proteome analysis and glycogen quantitation of field *Trichodesmium* populations.** *Trichodesmium* cells were sampled over a diel cycle on 10 March 2018 (Tricolim/AT39-05 expedition) from the subtropical Eastern Atlantic (65° 22.420' W, 17° 0.284' N). Samples of mixed morphology were collected in biological triplicates by gentle hand-picking, followed by two rinses in 0.2 µm sterile filtered trace-metal-clean seawater and decanting onto a 0.2 µm supor filter. Samples were flash frozen in liquid nitrogen and stored at -80 °C until analysis. Proteins were digested using the tube-gel method described above and analysed on a Michrom Advance HPLC coupled to a Q-exactive mass spectrometer (Thermo Fisher) with a Michrom Advance CaptiveSpray source. A 4 h one-dimension chromatography gradient (0.1% formic acid and 0.1% formic acid in 99.9% acetonitrile) was performed using a C18 column (0.1 × 150 mm ID, 3 µm particle size, 120 Å pore size, C18 Repronil-Gold, Maisch, packed in a New Objective PicoFrit column). MS1 scans were monitored between  $m/z$  380 and 1,280 with 5 s dynamic exclusion. Raw spectra were searched using the SequestHT algorithm implemented in Proteome Discoverer 2.2 using a publicly available *Trichodesmium* metagenome (JGI IMG ID 2821474806). Search parameters were 10 ppm parent mass tolerance, 0.6 Da fragment mass tolerance, cysteine modification +57.022 and methionine modification +16. Protein identifications were made via PeptideProphet implemented in Scaffold (Proteome Software) with 0.06% peptide and 1.5% protein global false discovery rates. In stations around this region, the population was dominated by a *Trichodesmium theibautii* species<sup>89</sup>. A total of 1,590 *Trichodesmium* proteins were identified.

For glycogen measurements, cells were lysed by boiling the filter split at 95 °C for 10 min. Lysate was clarified by centrifugation and glycogen was then quantified by the Sigma-Aldrich glycogen assay kit (MAK016) used in colorimetric mode. A separate set of samples collected on 7 August 2017 (JC150 expedition) from the Western Atlantic (31° W, 22° N) was also analysed for glycogen content<sup>42</sup>. These were collected in triplicate from early morning plankton nets deployed to different depths (40, 90 and 160 m), dragged for 10 min and then returned quickly to the surface, and glycogen content analysed as before.

**Estimation of terminal velocity of phytoplankton particles.** The terminal velocity of phytoplankton particles was calculated by solving for a terminal velocity as the solution to a force balance of buoyancy versus drag for representative spherical particles of diameters and specific gravities determined from the literature. The associated Matlab code is provided at [https://github.com/naheld/Held2020\\_TrichoDiel](https://github.com/naheld/Held2020_TrichoDiel). Particle properties are summarized in Supplementary Table 9. *Trichodesmium* specific gravity was calculated by allowing 25% of the cell volume to be occupied by a low-density gas vesicle (density = 1.025 kg m<sup>-3</sup>) and the remaining 75% of the cell volume to be occupied by *Trichodesmium* biomass of medium density as determined from the literature<sup>62,90</sup>.

The buoyancy force  $F_b$  was calculated from the particle's displacement and the density differential between the particle and surface ocean seawater (1,025 kg m<sup>-3</sup>) via Archimedes' principle:

$$F_b = \frac{4}{3}\pi r^3(\rho_T - \rho_{SW})g \quad (1)$$

where  $g$  denotes gravitational acceleration (9.81 m s<sup>-2</sup>),  $r$  denotes particle radius, and  $\rho_T$  and  $\rho_{SW}$  denote the densities of the phytoplankton and seawater, respectively. Positive  $F_b$  indicates a negatively buoyant (sinking) particle. The drag force was calculated according to

$$F_d = C_D(\text{Re})\frac{1}{2}\rho_{SW}v|\nu|\pi r^2 \quad (2)$$

where  $C_D(\text{Re})$  denotes the drag coefficient for a sphere referenced to cross-sectional area and depends on the Reynolds Number  $\text{Re}$ , and  $\nu$  denotes the terminal velocity of the particle. Positive values of  $\nu$  denote a sinking particle. Terminal velocity was computed as the solution to  $F_b = F_d$ . To permit a numerical solution, we employed an analytical approximation<sup>91</sup> to the drag coefficient as a function of  $\text{Re}$  accurate for smooth spheres to  $\text{Re} < 10^6$ . The Reynolds numbers corresponding to the terminal velocities determined for the range of particle sizes and velocities examined span many orders of magnitude, from laminar to turbulent flow regimes ( $\text{Re}$  range:  $4.4 \times 10^{-11} - 2.5 \times 10^3$ ).

For *Trichodesmium* colonies with hypothetical night nitrogen fixation, the density of the particle was increased by 60% (Table 1). The mass of dust particles was estimated from a previously published synchrotron-based X-ray fluorescence

element map of a natural *Trichodesmium* puff-type colony of diameter ~1 mm that contained dust particles (Extended Data Fig. 10)<sup>89</sup>. The integrated iron mass was calculated by isolating particles from the element map using a concentration mask and pixel density to estimate the integrated iron concentration of all particles in the colony. The estimated iron mass was 1.03 µg or 18.4 nmol Fe.

**Statistical analyses.** Downstream analyses including data preparation, plotting and statistical analyses were conducted in Python 3.0 using the matplotlib 3.3.4 (<http://matplotlib.org/>)<sup>92</sup>, seaborn 0.11.1 (<https://seaborn.pydata.org>) <https://doi.org/10.5281/zenodo.592845>) and scipy 1.6.1 (<https://docs.scipy.org/>)<sup>93</sup> libraries. Hierarchical clustering as in Fig. 1a,c was performed by the unweighted pair group method with arithmetic mean (UPGMA) implemented in scipy and plotted in seaborn. Sinusoidal modelling for Fig. 1e,f was performed by optimization of a basic sinusoidal function via least-squares regression implemented in scipy. Protein expression networks were generated with the Python networkX 2.4 library ([https://networkx.github.io](https://networkx.github.io/))<sup>94</sup> using pre-calculated Spearman correlation statistics (Supplementary Table 8), where a positive correlation between two proteins was defined as Spearman correlation coefficient >0.8,  $P < 0.05$  (two-sided Student's  $t$ -test), and a negative correlation between two proteins defined as Spearman correlation coefficient < -0.8,  $P < 0.05$  (two-sided Student's  $t$ -test).

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The mass spectrometry proteomics data has been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifiers PXD016332 and <https://doi.org/10.6019/PXD016332> (laboratory experiments) and identifier PXD027796 and <https://doi.org/10.6019/PXD027796> (field data). The processed proteomic data are also available at the Biological and Chemical Oceanography Data Management Office (BCO-DMO) (<https://www.bco-dmo.org/dataset/783873>). Source data are provided for main text Figs. 1–5 and Extended Data Figs. 1–10. Source data are provided with this paper.

## Code availability

Fully reproducible code for sinking velocity calculations, statistics and plotting is available at [https://github.com/naheld/Held2020\\_TrichoDiel](https://github.com/naheld/Held2020_TrichoDiel).

Received: 1 March 2021; Accepted: 17 November 2021;

Published online: 10 January 2022

## References

- Zehr, J. P. & Capone, D. G. Changing perspectives in marine nitrogen fixation. *Science* **9514**, 729 (2020).
- Karl, D. et al. Dinitrogen fixation in the world's oceans. *Biogeochemistry* **57–58**, 47–98 (2002).
- Dugdale, R. & Wilkerson, F. in *Primary Productivity and Biogeochemical Cycles in the Sea* (eds Falkowski, P. G. et al.) 107–122 (Springer, 1992).
- Carpenter, E. J. & Capone, D. G. in *Nitrogen in the Marine Environment* 2nd edn (eds Capone, D. G., Bronk, D. A., Mulholland, M. R. & Carpenter, E. J.) Ch. 4 (Elsevier, 2008).
- Gruber, N. & Sarmiento, J. L. Global patterns of marine nitrogen fixation and denitrification. *Global Biogeochem. Cycles* **11**, 23–266 (1997).
- Buchanan, P. J., Chase, Z., Matear, R. J., Phipps, S. J. & Bindoff, N. L. Marine nitrogen fixers mediate a low latitude pathway for atmospheric CO<sub>2</sub> drawdown. *Nat. Commun.* <https://doi.org/10.1038/s41467-019-12549-z> (2019).
- Monteiro, F. M., Follows, M. J. & Dutkiewicz, S. Distribution of diverse nitrogen fixers in the global ocean. *Global Biogeochem. Cycles* **24**, 1–16 (2010).
- Church, M. J., Björkman, K. M., Karl, D. M., Saito, M. A. & Zehr, J. P. Regional distributions of nitrogen-fixing bacteria in the Pacific Ocean. *Limnol. Oceanogr.* **53**, 63–77 (2008).
- Monteiro, F. M., Dutkiewicz, S. & Follows, M. J. Biogeographical controls on the marine nitrogen fixers. *Global Biogeochem. Cycles* **25**, 1–8 (2011).
- Dutkiewicz, S., Ward, B. A., Monteiro, F. & Follows, M. J. Interconnection of nitrogen fixers and iron in the Pacific Ocean: theory and numerical simulations. *Global Biogeochem. Cycles* **26**, 1–16 (2012).
- Walworth, N. G. et al. Nutrient-colimited *Trichodesmium* as a nitrogen source or sink in a future ocean. *Appl. Environ. Microbiol.* **84**, 1–14 (2018).
- McGillicuddy, D. J. Jr. Do *Trichodesmium* spp. populations in the North Atlantic export most of the nitrogen they fix? *Global Biogeochem. Cycles* **28**, 103–114 (2014).
- Carpenter, E. J. & Romans, K. Major role of the cyanobacterium *Trichodesmium* in nutrient cycling in the North Atlantic Ocean. *Science* **254**, 1989–1992 (1991).
- Bergman, B., Sandh, G., Lin, S., Larsson, J. & Carpenter, E. J. *Trichodesmium* – a widespread marine cyanobacterium with unusual nitrogen fixation properties. *FEMS Microbiol. Rev.* **37**, 286–302 (2013).

15. Capone, D. G. *Trichodesmium*, a globally significant marine cyanobacterium. *Science* **276**, 1221–1229 (1997).
16. Gallon, J. R. The oxygen sensitivity of nitrogenase: a problem for biochemists and micro-organisms. *Trends Biochem. Sci.* **6**, 19–23 (1981).
17. Saito, M. A. et al. Iron conservation by reduction of metalloenzyme inventories in the marine diazotroph *Crocospaera watsonii*. *Proc. Natl Acad. Sci. USA* **108**, 2184–2189 (2011).
18. Dron, A. et al. Light-dark (12:12) cycle of carbon and nitrogen metabolism in *Crocospaera watsonii* WH8501: relation to the cell cycle. *Environ. Microbiol.* **14**, 967–981 (2012).
19. Mohr, W., Intermaggio, M. P. & LaRoche, J. Diel rhythm of nitrogen and carbon metabolism in the unicellular, diazotrophic cyanobacterium *Crocospaera watsonii* WH8501. *Environ. Microbiol.* **12**, 412–421 (2010).
20. Flores, E. & Herrero, A. Compartmentalized function through cell differentiation in filamentous cyanobacteria. *Nat. Rev. Microbiol.* **8**, 39–50 (2010).
21. Burnat, M., Herrero, A. & Flores, E. Compartmentalized cyanophycin metabolism in the diazotrophic filaments of a heterocyst-forming cyanobacterium. *Proc. Natl Acad. Sci. USA* **111**, 3823–3828 (2014).
22. Sherman, D. M., Tucker, D. & Sherman, L. A. Heterocyst development and localization of cyanophycin in  $N_2$ -fixing cultures of *Anabaena* sp. PCC 7120 (Cyanobacteria). *J. Phycol.* **94**, 932–941 (2000).
23. Lamont, H. C., Silvester, W. B. & Torrey, J. G. Nile red fluorescence demonstrates lipid in the envelope of vesicles from  $N_2$ -fixing cultures of *Frankia*. *Can. J. Microbiol.* **34**, 656–660 (1988).
24. Saino, T. Diel variation in nitrogen fixation by a marine blue-green alga, *Trichodesmium thiebautii*. *Deep Sea Res.* **25**, 1259–1263 (1978).
25. Saino, T. & Hattori, A. Aerobic nitrogen fixation by the marine non-heterocystous cyanobacterium *Trichodesmium* (*Oscillatoria*) spp.: its protective mechanism against oxygen. *Mar. Biol.* **70**, 251–254 (1982).
26. Berman-Frank, I. et al. Segregation of nitrogen fixation and oxygenic photosynthesis in the marine cyanobacterium *Trichodesmium*. *Science* **294**, 1534–1537 (2001).
27. Ohki, K. & Taniuchi, Y. Detection of nitrogenase in individual cells of a natural population of *Trichodesmium* using immunocytochemical methods for fluorescent cells. *J. Oceanogr.* **65**, 427–432 (2009).
28. Eichner, M. et al.  $N_2$  fixation in free-floating filaments of *Trichodesmium* is higher than in transiently suboxic colony microenvironments. *New Phytol.* **222**, 852–863 (2019).
29. Ohki, K. Intercellular localization of nitrogenase in a non-heterocystous cyanobacterium (cyanophyte), *Trichodesmium* sp. NIBB1067. *J. Oceanogr.* **64**, 211–216 (2008).
30. Ohki, K., Zehr, F. & Fujita, Y. Regulation of nitrogenase activity in relation to the light-dark regime in the filamentous non-heterocystous cyanobacterium *Trichodesmium* sp. NIBB 1067. *J. Gen. Microbiol.* **138**, 2679–2685 (1992).
31. Finzi-Hart, J. A. et al. Fixation and fate of C and N in the cyanobacterium *Trichodesmium* using nanometer-scale secondary ion mass spectrometry. *Proc. Natl Acad. Sci. USA* **106**, 9931 (2009).
32. Sandh, G., El-Shehawry, R., Diez, B. & Bergman, B. Temporal separation of cell division and diazotrophy in the marine diazotrophic cyanobacterium *Trichodesmium erythraeum* IMS101. *FEMS Microbiol. Lett.* **295**, 281–288 (2009).
33. Küpper, H. et al. Traffic lights in *Trichodesmium*. Regulation of photosynthesis for nitrogen fixation studied by chlorophyll fluorescence kinetic microscopy. *Plant Physiol.* **135**, 2120–2133 (2019).
34. Ohki, K. & Fujita, Y. Aerobic nitrogenase activity measured as acetylene reduction in the marine non-heterocystous cyanobacterium *Trichodesmium* spp. grown under artificial conditions. *Mar. Biol.* **98**, 111–114 (1988).
35. Waterbury, J. B. & Willey, J. M. Isolation and growth of marine planktonic Cyanobacteria. *Methods Enzymol.* **167**, 100–105 (1988).
36. Chen, Y. B., Zehr, J. P. & Mellon, M. Growth and nitrogen fixation of the diazotrophic filamentous nonheterocystous cyanobacterium *Trichodesmium* sp. IMS 101 in defined media: evidence for a circadian rhythm. *J. Phycol.* **32**, 916–923 (1996).
37. Berman-Frank, I., Bidle, K. D., Haramaty, L. & Falkowski, P. G. The demise of the marine cyanobacterium, *Trichodesmium* spp., via an autocatalyzed cell death pathway. *Limnol. Oceanogr.* **49**, 997–1005 (2004).
38. Bell, P. R. F. et al. Laboratory culture studies of *Trichodesmium* isolated from the Great Barrier Reef lagoon, Australia. *Hydrobiologia* **532**, 9–21 (2005).
39. Tzubari, Y., Magnezi, L., Be'Er, A. & Berman-Frank, I. Iron and phosphorus deprivation induce sociality in the marine bloom-forming cyanobacterium *Trichodesmium*. *ISME J.* **12**, 1682–1693 (2018).
40. Held, N. A., McIlvin, M. R., Moran, D. M., Laub, M. T. & Saito, M. A. Unique patterns and biogeochemical relevance of two-component sensing in marine bacteria. *mSystems* **4**, 1–16 (2019).
41. Aryal, U. K. & Sherman, L. A. in *Cyanobacteria Omics Manipulation* (ed. Los, D. A.) Ch. 6 (Caister Academic Press, 2017).
42. Held, N. A. et al. Co-occurrence of Fe and P stress in natural populations of the marine diazotroph *Trichodesmium*. *Biogeosciences* **17**, 2537–2551 (2020).
43. Klugkist, J., Haaker, H., Wassink, H. & Veeger, C. The catalytic activity of nitrogenase in intact *Azotobacter vinelandii* cells. *Eur. J. Biochem.* **146**, 509–515 (1985).
44. Zehr, J. P., Wyman, M., Miller, V., Capone, D. G. & Duguay, L. Modification of the Fe protein of nitrogenase in natural populations of *Trichodesmium thiebautii*. *Appl. Environ. Microbiol.* **59**, 669–676 (1993).
45. Rodriguez, I. B. & Ho, T.-Y. Diel nitrogen fixation pattern of *Trichodesmium*: the interactive control of light and Ni. *Sci. Rep.* **4**, 4445 (2014).
46. Eichner, M., Kranz, S. A. & Rost, B. Combined effects of different  $CO_2$  levels and N sources on the diazotrophic cyanobacterium *Trichodesmium*. *Physiol. Plant.* **152**, 316–330 (2014).
47. Hutchins, D. A. et al. Irreversibly increased nitrogen fixation in *Trichodesmium* experimentally adapted to elevated carbon dioxide. *Nat. Commun.* **6**, 1–7 (2015).
48. Levitan, O. et al. Combined effects of  $CO_2$  and light on the  $N_2$ -fixing cyanobacterium *Trichodesmium* IMS101: a mechanistic view. *Plant Physiol.* **154**, 346–356 (2010).
49. Villareal, T. A. & Carpenter, E. J. Buoyancy regulation and the potential for vertical migration in the oceanic cyanobacterium *Trichodesmium*. *Microb. Ecol.* **45**, 1–10 (2003).
50. Rabouille, S., Staal, M., Stal, L. J. & Soetaert, K. Modeling the dynamic regulation of nitrogen fixation in the cyanobacterium *Trichodesmium* sp. *Appl. Environ. Microbiol.* **72**, 3217–3227 (2006).
51. Breitbarth, E., Wohlers, J., Kläs, J., LaRoche, J. & Peeken, I. Nitrogen fixation and growth rates of *Trichodesmium* IMS-101 as a function of light intensity. *Mar. Ecol. Prog. Ser.* **359**, 25–36 (2008).
52. Chen, Y. B. et al. Circadian rhythm of nitrogenase gene expression in the diazotrophic filamentous nonheterocystous cyanobacterium *Trichodesmium* sp. strain IMS101. *J. Bacteriol.* **180**, 3598–3605 (1998).
53. Rabouille, S., Staal, M., Stal, L. J. & Soetaert, K. Modeling the dynamic regulation of nitrogen fixation in the *Cyanobacterium Trichodesmium* sp. *Appl. Environ. Microbiol.* **72**, 3217–3227 (2006).
54. Capone, D. G., O'Neill, J. M., Zehr, J. & Carpenter, E. J. Basis for diel variation in nitrogenase activity in the marine planktonic cyanobacterium *Trichodesmium thiebautii*. *Appl. Environ. Microbiol.* **56**, 3532–3536 (1990).
55. Gründel, M., Scheunemann, R., Lockau, W. & Zilliges, Y. Impaired glycogen synthesis causes metabolic overflow reactions and affects stress responses in the cyanobacterium *Synechocystis* sp. PCC 6803. *Microbiology* **158**, 3032–3043 (2012).
56. Jackson, S. A., Eaton-Rye, J. J., Bryant, D. A., Posewitz, M. C. & Davies, F. K. Dynamics of photosynthesis in a glycogen-deficient glgC mutant of *Synechococcus* sp. strain PCC 7002. *Appl. Environ. Microbiol.* **81**, 6210–6222 (2015).
57. Boatman, T. G., Davey, P. A., Lawson, T. & Geider, R. J. The physiological cost of diazotrophy for *Trichodesmium erythraeum* IMS101. *PLoS ONE* **13**, 1–24 (2018).
58. Chappell, P. D., Moffett, J. W., Hynes, A. M. & Webb, E. A. Molecular evidence of iron limitation and availability in the global diazotroph *Trichodesmium*. *ISME J.* **6**, 1728–1739 (2012).
59. Chappell, P. D. & Webb, E. A. A molecular assessment of the iron stress response in the two phylogenetic clades of *Trichodesmium*. *Environ. Microbiol.* **12**, 13–27 (2010).
60. Walsby, A. E. The properties and buoyancy-providing role of gas vacuoles in *Trichodesmium* Ehrenberg. *Br. Phycol. J.* **13**, 103–116 (1978).
61. Villareal, T. A. & Carpenter, E. J. Diel buoyancy regulation in the marine diazotrophic cyanobacterium *Trichodesmium thiebautii*. *Limnol. Oceanogr.* **35**, 1832–1837 (1990).
62. Romans, K. M., Carpenter, E. J. & Bergman, B. Buoyancy regulation in the colonial diazotrophic cyanobacterium *Trichodesmium tenue*: ultrastructure and storage of carbohydrate, polyphosphate, and nitrogen. *J. Phycol.* **30**, 935–942 (1994).
63. Wang, L. et al. Molecular structure of glycogen in *Escherichia coli*. *Biomacromolecules* **20**, 2821–2829 (2019).
64. Berman-Frank, I., Cullen, J. T., Shaked, Y., Sherrell, R. M. & Falkowski, P. G. Iron availability, cellular iron quotas, and nitrogen fixation in *Trichodesmium*. *Limnol. Oceanogr.* **46**, 1249–1260 (2001).
65. Kustka, A. B. et al. Iron requirements for dinitrogen- and ammonium-supported growth in cultures of *Trichodesmium* (IMS 101): comparison with nitrogen fixation rates and iron:carbon ratios of field populations. *Limnol. Oceanogr.* **49**, 1224 (2004).
66. Paerl, H. W., Prufert-Bebout, I. L. E., Guo, C. & Carolina, N. Iron-stimulated  $N_2$  fixation and growth in natural and cultured populations of the planktonic marine cyanobacteria *Trichodesmium* spp. *Appl. Environ. Microbiol.* **60**, 1044–1047 (1994).
67. Rubin, M., Berman-Frank, I. & Shaked, Y. Dust- and mineral-iron utilization by the marine dinitrogen-fixer *Trichodesmium*. *Nat. Geosci.* **4**, 529–534 (2011).
68. Polyviou, D. et al. Desert dust as a source of iron to the globally important diazotroph *Trichodesmium*. *Front. Microbiol.* **8**, 1–12 (2018).

69. Basu, S. & Shaked, Y. Mineral iron utilization by natural and cultured *Trichodesmium* and associated bacteria. *Limnol. Oceanogr.* **63**, 2307–2320 (2018).
70. Held, N. A. et al. Mechanisms and heterogeneity of in situ mineral processing by the marine nitrogen fixer *Trichodesmium* revealed by single-colony metaproteomics. *ISME Commun.* **1**, 35 (2021).
71. Basu, S., Gledhill, M., de Beer, D., Prabhu Matondkar, S. G. & Shaked, Y. Colonies of marine cyanobacteria *Trichodesmium* interact with associated bacteria to acquire iron from dust. *Commun. Biol.* **2**, 1–8 (2019).
72. Tyrrell, T. et al. Large-scale latitudinal distribution of *Trichodesmium* spp. in the Atlantic Ocean. *J. Plankton Res.* **25**, 405–416 (2003).
73. Robson, R. L. & Postgate, J. R. Oxygen and hydrogen in biological nitrogen fixation. *Annu. Rev. Microbiol.* **34**, 183–207 (1980).
74. Zehr, J. P. Nitrogen fixation by marine cyanobacteria. *Trends Microbiol.* **19**, 162–173 (2011).
75. Bergman, B. & Carpenter, E. J. Nitrogenase confined to randomly distributed trichomes in the marine cyanobacterium *Trichodesmium thiebautii*. *J. Phycol.* **27**, 158–165 (1991).
76. Inomura, K., Wilson, S. T. & Deutsch, C. Mechanistic model for the coexistence of nitrogen fixation and photosynthesis in marine *Trichodesmium*. *mSystems* **4**, 1–13 (2019).
77. Janson, S., Matveyev, A. & Bergman, B. The presence and expression of hetR in the non-heterocystous cyanobacterium *Symploca* PCC 8002. *FEMS Microbiol. Lett.* **168**, 173–179 (1998).
78. Zhang, J. Y., Chen, W. L. & Zhang, C. C. hetR and patS, two genes necessary for heterocyst pattern formation, are widespread in filamentous nonheterocyst-forming cyanobacteria. *Microbiology* **155**, 1418–1426 (2009).
79. Moore, J. K., Doney, S. C., Glover, D. M. & Fung, I. Y. Iron cycling and nutrient-limitation patterns in surface waters of the world ocean. *Deep Sea Res. 2 Top. Stud. Oceanogr.* **49**, 463–507 (2001).
80. Chisholm, S. W. in *Primary Productivity and Biogeochemical Cycles in the Sea* (eds Falkowski, P. G. et al.) 213–237 (Springer, 1992). [https://doi.org/10.1007/978-1-4899-0762-2\\_12](https://doi.org/10.1007/978-1-4899-0762-2_12)
81. Young, K. D. The selective value of bacterial shape. *Microbiol. Mol. Biol. Rev.* **70**, 660–703 (2006).
82. Lu, X. & Zhu, H. Tube-gel digestion: a novel proteomic approach for high-throughput analysis of membrane proteins. *Mol. Cell Proteom.* **4**, 1948–1958 (2005).
83. Saito, M. A. et al. Multiple nutrient stresses at intersecting Pacific Ocean biomes detected by protein biomarkers. *Science* **345**, 1173–1177 (2014).
84. McIlvin, M. R. & Saito, M. A. Online nanoflow two-dimension comprehensive active modulation reversed phase-reversed phase liquid chromatography high-resolution mass spectrometry for metaproteomics of environmental and microbiome samples. *J. Proteome Res.* **20**, 4589–4597 (2021).
85. Lee, M. D. et al. Transcriptional activities of the microbial consortium living with the marine nitrogen-fixing cyanobacterium *Trichodesmium* reveal potential roles in community-level nitrogen cycling. *Appl. Environ. Microbiol.* **84**, AEM.02026-17 (2017).
86. Zhang, Y., Wen, Z., Washburn, M. P. & Florens, L. Refinements to label-free proteome quantitation: how to deal with peptides shared by multiple proteins. *Anal. Chem.* **82**, 2272–2281 (2010).
87. Gallien, S., Bourmaud, A., Kim, S. Y. & Domon, B. Technical considerations for large-scale parallel reaction monitoring analysis. *J. Proteom.* **100**, 147–159 (2014).
88. Pino, L. K. et al. The skyline ecosystem: informatics for quantitative mass spectrometry proteomics. *Mass Spectrom. Rev.* **176**, 139–148 (2019).
89. Held, N. A. et al. Mechanisms and heterogeneity of in situ mineral processing by the marine nitrogen fixer *Trichodesmium* revealed by single-colony metaproteomics. *ISME Commun.* <https://doi.org/10.1038/s43705-021-00034-y> (2021).
90. White, A. E., Spitz, Y. H. & Letelier, R. M. Modeling carbohydrate ballasting by *Trichodesmium* spp. *Mar. Ecol. Prog. Ser.* **323**, 35–45 (2006).
91. Morrison, F. A. *An Introduction to Fluid Mechanics* (Cambridge Univ. Press, 2013).
92. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
93. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
94. Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring network structure, dynamics, and function using NetworkX. In *7th Python Scientific Conference (SciPy 2008)* 11–15 (2008).

## Acknowledgements

This work was supported by NSF Graduate Research Fellowship grant 1122274 (N.A.H.), Gordon and Betty Moore Foundation grant GBMF-3782 (M.A.S.), National Science Foundation grants OCE-1657766, OCE-1850719 and OCE-1924554 (M.A.S.), National Institutes of Health grant GM135709-02 (M.A.S.), and the Woods Hole Oceanographic Institution Ocean Ventures Fund (N.A.H.). N.A.H. was additionally supported by Principles of Microbial Ecosystems collaboration of the Simons Foundation (grant ID 542379). We thank the scientific staff and crew of the AT39-05/Tricolim research expedition, particularly chief scientist D. Hutchins, and the JC150/Trizloc expedition, particularly chief scientist C. Mahaffey. Special thanks to B. White.

## Author contributions

N.A.H., J.B.W. and M.A.S. conceived and designed the experiment. N.A.H., R.M.K., D.M.M. and F.W.V. performed the experiments and analyses. M.R.M. and N.A.H. performed mass spectrometry analyses. M.J., N.A.H. and M.A.S. developed sinking velocity calculations. K.M.S. performed synchrotron element map analyses. All authors contributed to the writing and editing of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41564-021-01028-1>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41564-021-01028-1>.

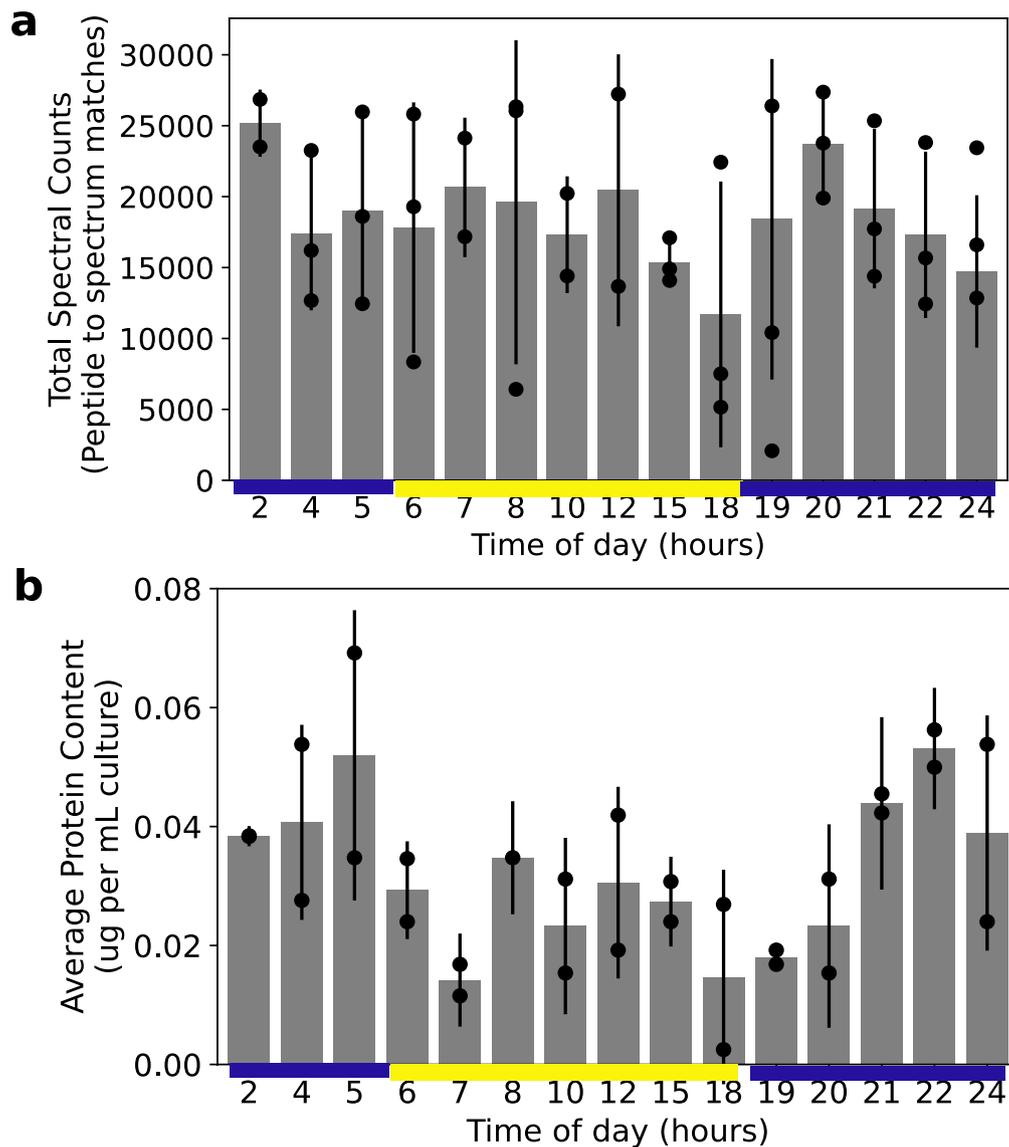
**Correspondence and requests for materials** should be addressed to Mak A. Saito.

**Peer review information** *Nature Microbiology* thanks Ilka Axmann, David Karl, Adam Kustka and Marc Strous for their contribution to the peer review of this work.

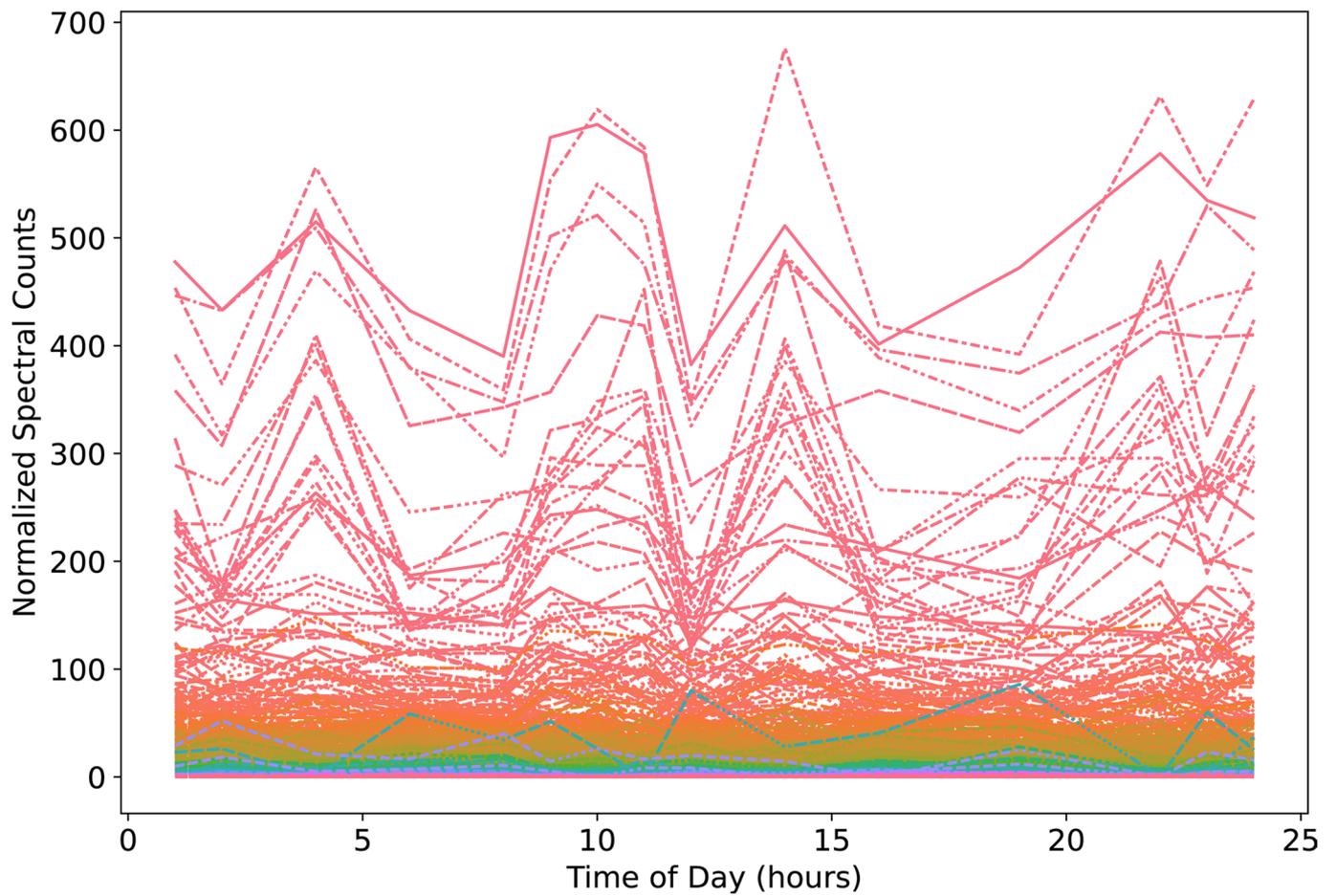
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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

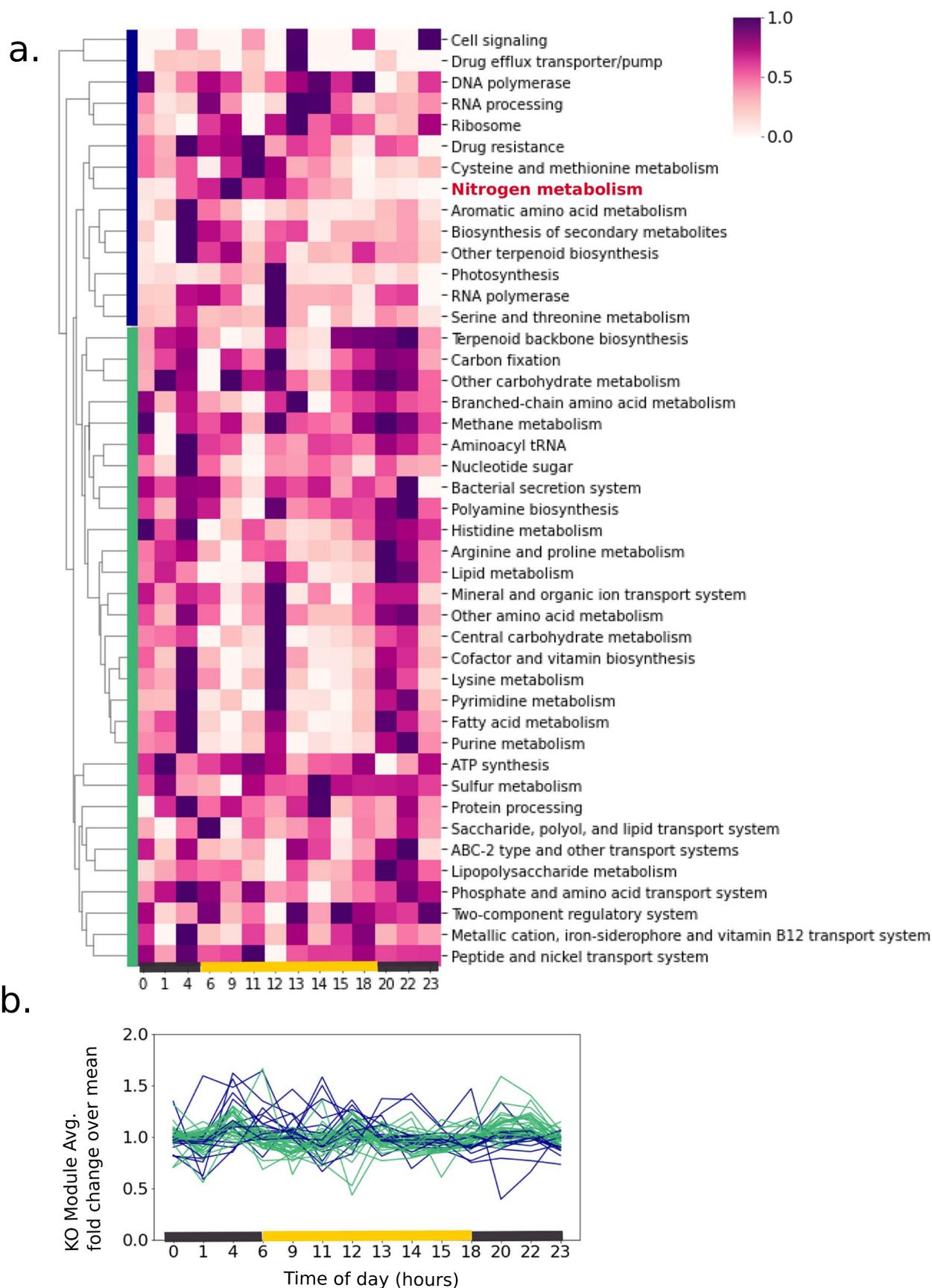
© The Author(s), under exclusive licence to Springer Nature Limited 2022



**Extended Data Fig. 1 | Total spectral counts and average protein content across the diel cycle. a)** Average total spectral counts (peptide to spectrum matches) with error bar representing  $\pm$  one standard deviation, at each time point. Each data point is also shown individually as black scatter points. Yellow and indigo bars indicate the light and dark periods, respectively. Total spectral counts were relatively uniform and do not vary systematically throughout the diel cycle, implying consistency in the proteome analyses. **b)** Total protein content in the culture shown with error bar representing  $\pm$  one standard deviation, for biological duplicates after protein precipitation and purification, measured by a colorimetric assay. Higher protein abundances at night may suggest nighttime cell growth. Again, each data point is also shown individually as black scatter points. Yellow and indigo bars indicate the light and dark periods, respectively.

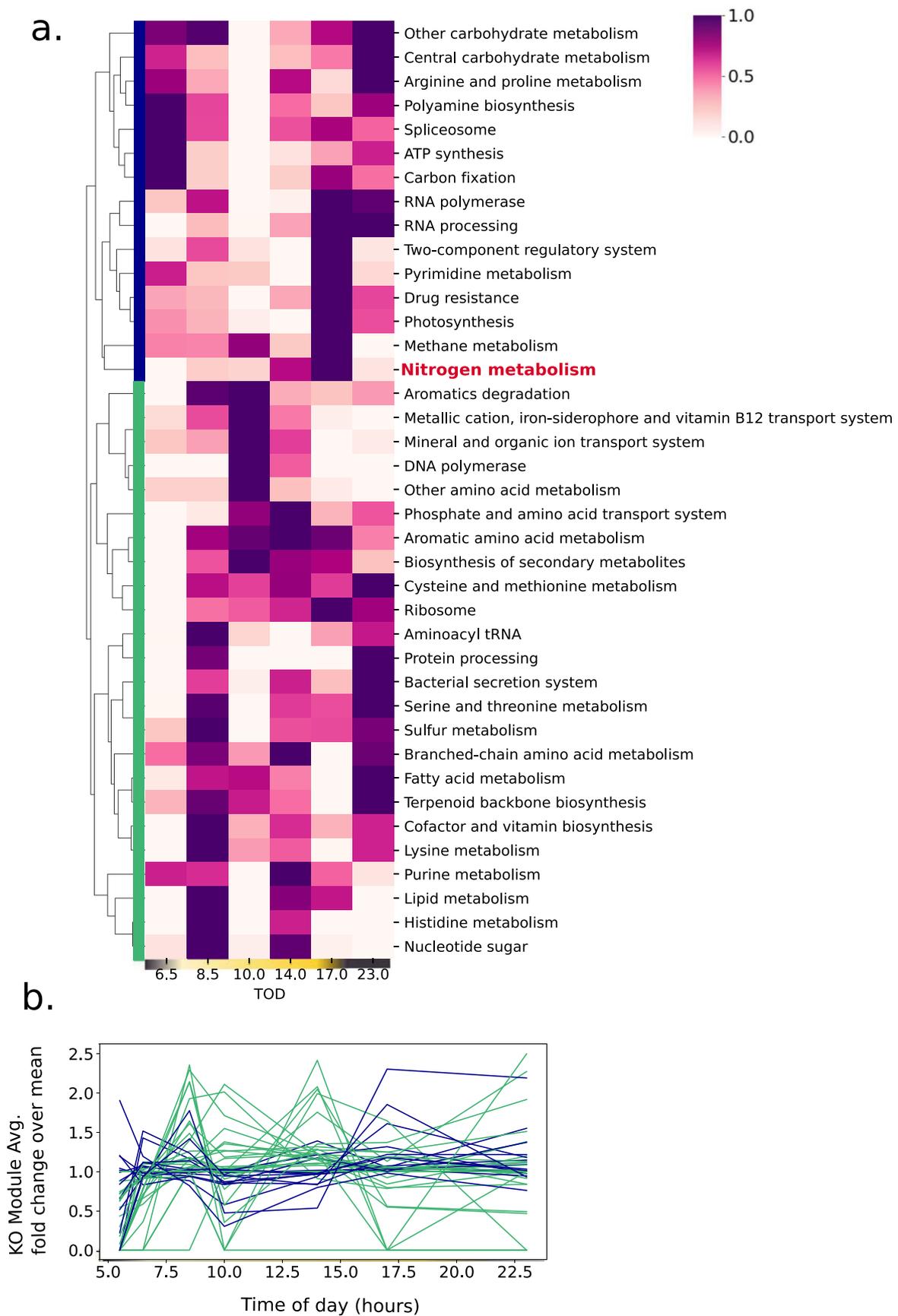


**Extended Data Fig. 2 | Dynamics of the entire proteome of *Trichodesmium erythraeum* sp. IMS101 over the diel cycle.** The dynamic range of the normalized spectral count data can be observed, as well as fluctuations in protein abundance occurring throughout the experiment.



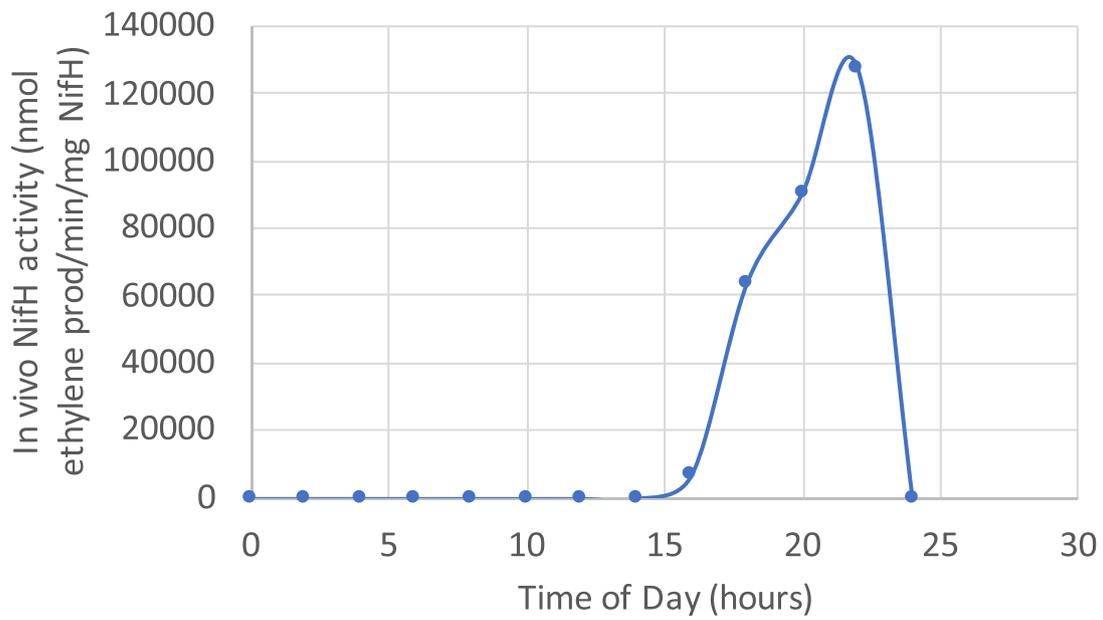
Extended Data Fig. 3 | See next page for caption.

**Extended Data Fig. 3 | Proteome dynamics of separate replicate laboratory experiment over the diel cycle. a)** Clustered heatmap of a replicate diel experiment conducted one year prior to the main experiment, under the same experimental conditions. Protein abundances were summed for each KO module and normalized across each row. **b)** Dynamics of the proteome clusters over the diel cycle, with each KO module represented as a line and colored based on the clustering in panel (A). Rapid oscillations of the proteome and clustering of the nitrogenase/nitrogen metabolism proteins with the photosystems are similar in the main experiment. Yellow and dark purple bars indicate the light and dark periods, respectively.

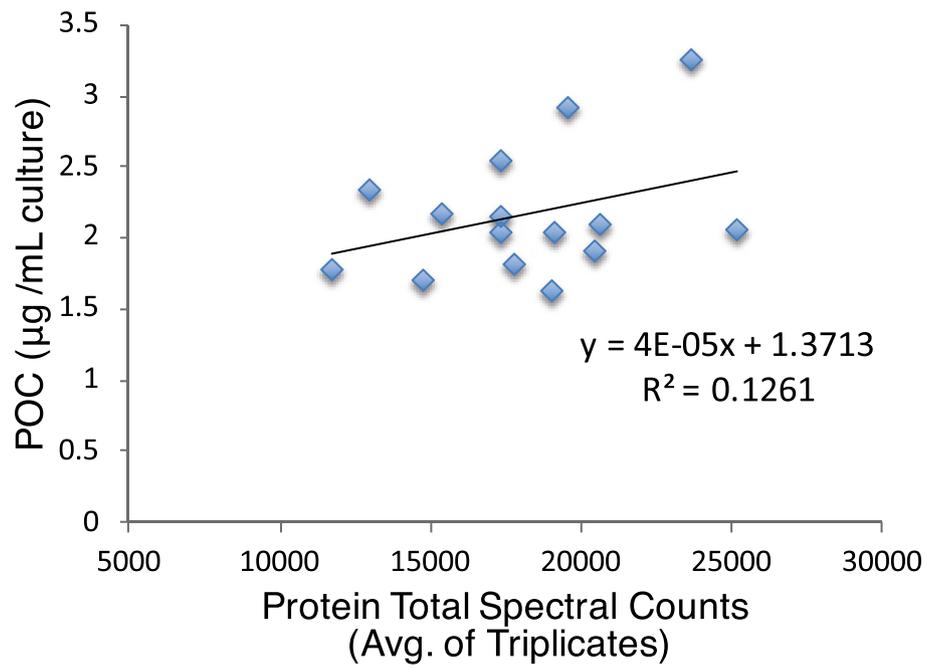


Extended Data Fig. 4 | See next page for caption.

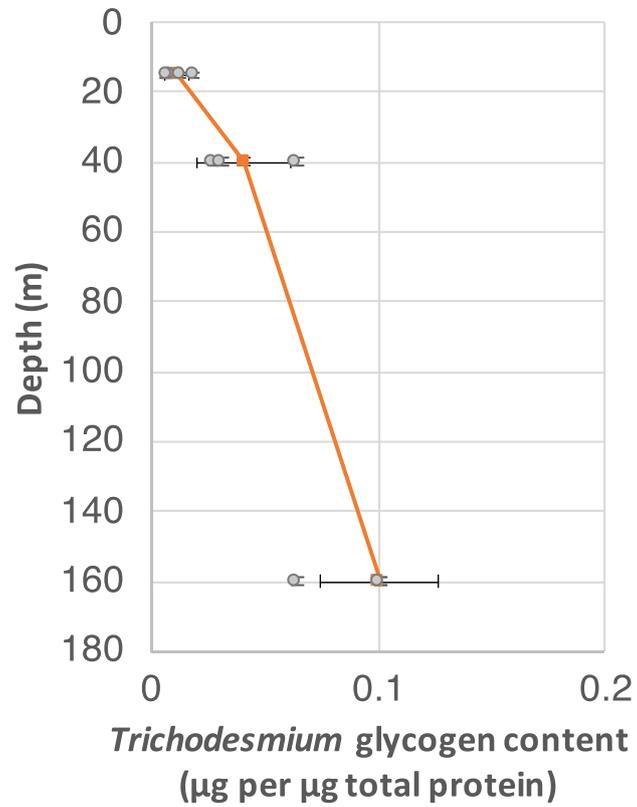
**Extended Data Fig. 4 | Proteome dynamics of field *Trichodesmium* population over the diel cycle.** **a)** Clustered heatmap of the proteome of a field *Trichodesmium* population sampled in situ over a diel cycle. Protein abundances were summed for each KO module and normalized across each row. **b)** Dynamics of the proteome clusters over the diel cycle, with each KO module represented as a line and colored based on the clustering in panel (A). Though the sampling was lower resolution than in the laboratory experiments, the rapid oscillations of the proteome are reproduced. Yellow and dark purple bars indicate the light and dark periods, respectively.



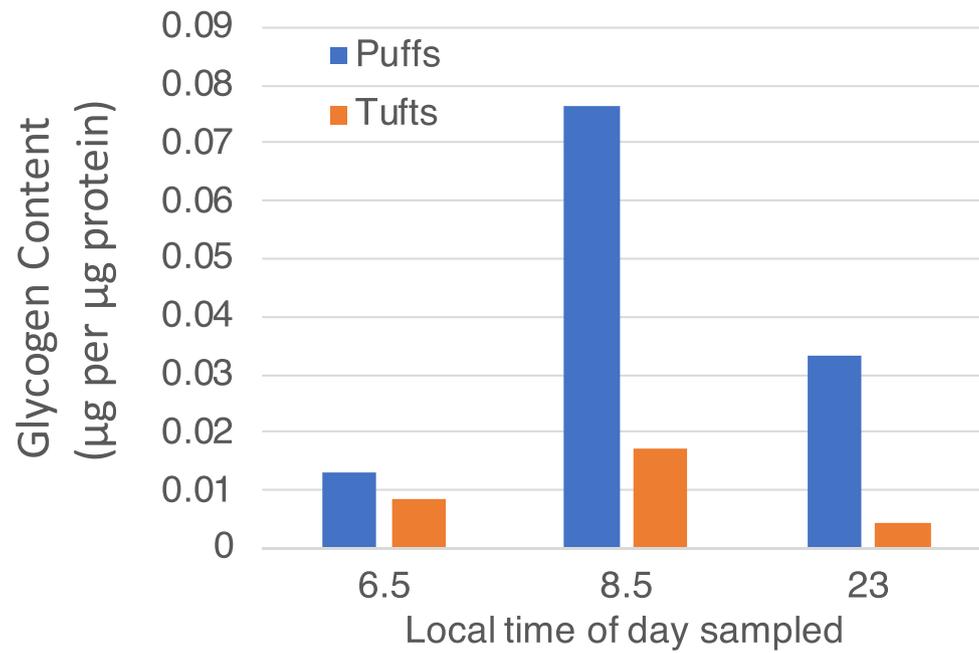
**Extended Data Fig. 5 | In vivo specific activity of the nitrogenase NifH protein over the diel cycle for *Crocosphaera watsonii*.** In vivo specific activity of the nitrogenase NifH protein (nmol ethylene produced per min per mg NifH) over the diel cycle for *Crocosphaera watsonii*<sup>34</sup>. Unlike in *Trichodesmium* which exhibits significant variability in nitrogenase activity throughout the diel cycle, in *Crocosphaera* nitrogenase is either not present or highly present and very active.



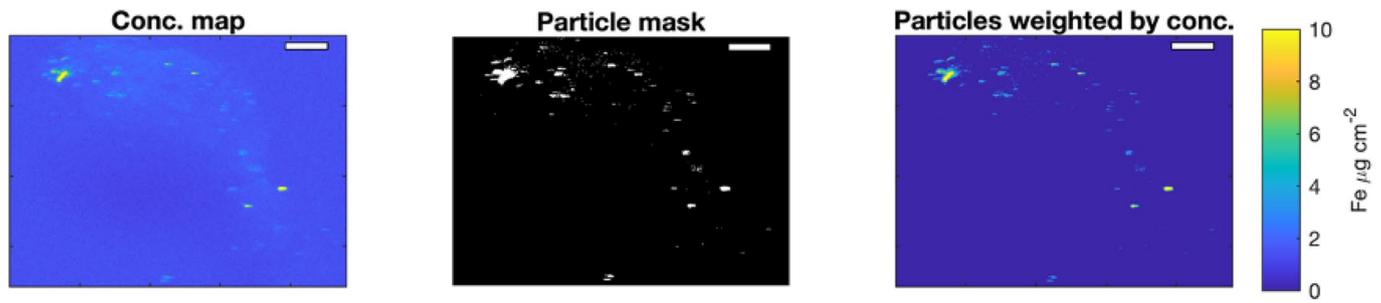
**Extended Data Fig. 6 | POC content versus total protein spectral counts.** POC content versus total protein spectral counts in the main laboratory experiment. These are weakly correlated suggesting that POC content is driven mainly by carbohydrate content, not protein abundance.



**Extended Data Fig. 7 | Glycogen content of *Trichodesmium* populations sampled in situ by depth.** The populations were sampled on August 7, 2017 at 31°W 22°N in the early morning. Error bars are standard deviations of the mean value of the biological triplicates, and corresponding data points are plotted in grey circles. For each depth,  $n = 3$  samples collected from replicate phytoplankton net sampling events,  $n = 2$  samples for depth = 160 m.

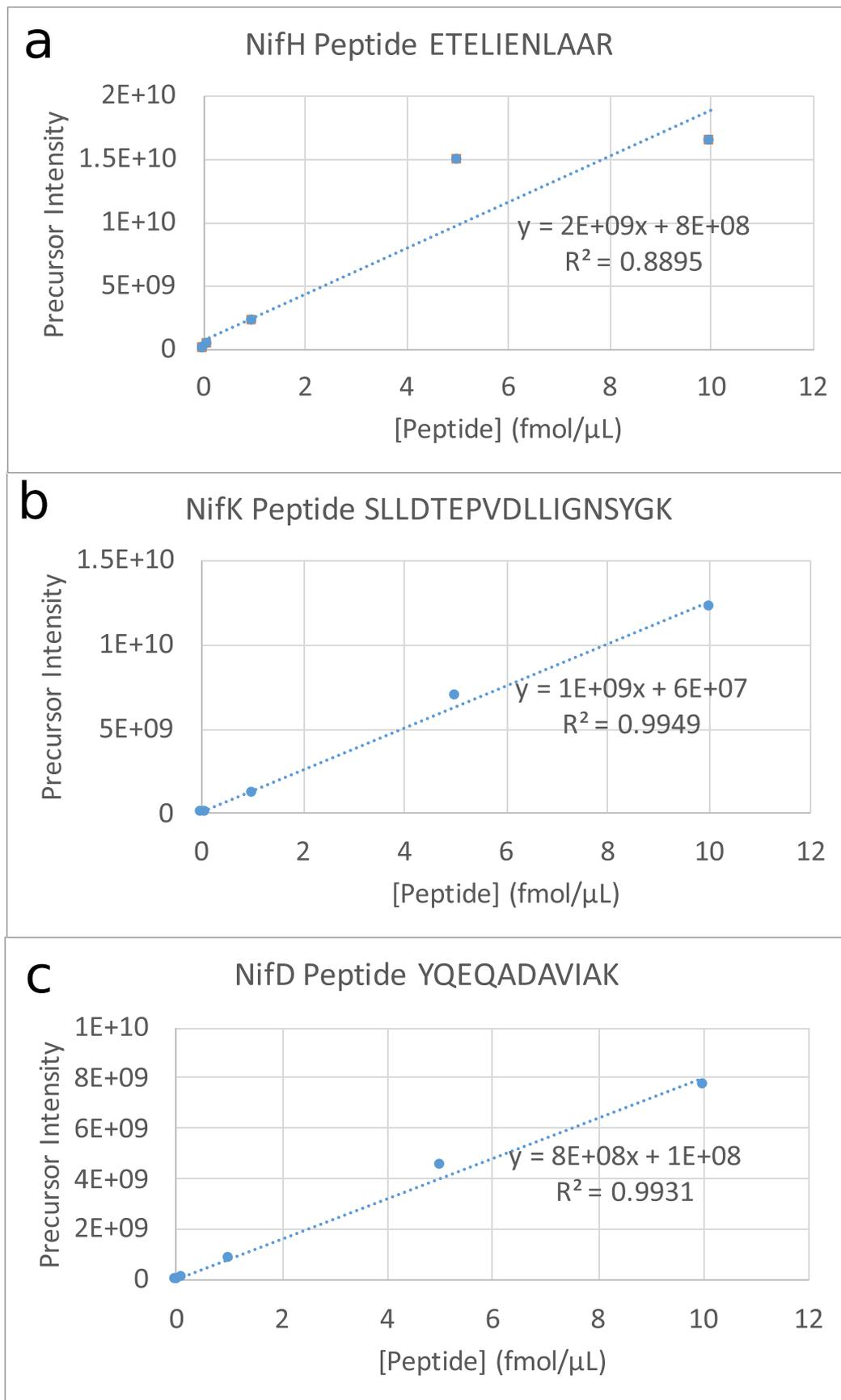


**Extended Data Fig. 8 | Glycogen content of *Trichodesmium* colonies separated by morphology.** Glycogen content of *Trichodesmium* colonies sampled in situ at the surface and separated by morphology. The populations were sampled from the surface on March 10, 2018 at 65 22.420 °W 17 0.284 °N and separated by morphology at the time of picking.



**Extended Data Fig. 9 | Synchrotron-based element maps used to determine mass of particulate iron associated with a puff-type colony.**

Synchrotron-based element maps used to determine mass of particulate iron associated with a puff-type colony, data originally collected as in Held et al., 2020<sup>20</sup>. The left image is the X-ray fluorescence-based concentration, the middle image represents pixels with sufficiently high Fe to be considered a particle, and the right image is the product of the left and middle images. The total particulate Fe was determined as the area integrated Fe of the right image. The scale bar represents 180 microns. As detailed in Held et al., 2021, five *Trichodesmium* colonies of differing morphologies and degrees of particle association were examined in this way. These images are representative of a *Trichodesmium* colony with average-to-high particle loading.



**Extended Data Fig. 10 |** Calibration curves for  $^{15}\text{N}$  labeled standard peptides used for absolute quantitation of the nitrogenase proteins. Precursor ion intensities were linearly correlated with analyzed peptide concentrations between 0-10 fmol  $\mu\text{L}^{-1}$ .

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | n/a                                 | Confirmed  |
|-------------------------------------|--|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of all covariates tested   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated   |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

**Data collection** Peptide to spectrum matching was performed using Sequest HT implemented in Proteome Discoverer v2.2. Protein identification and grouping and spectral count normalization was performed in Scaffold 4.8.7. Analysis of absolute protein concentrations from the PRM datasets was conducted in Skyline v20.2.

**Data analysis** Python 3.0, matplotlib 3.3.4, seaborn 0.11.1, scipy 1.6.1, networkX 2.4, matlab R2020a. Analysis and plotting scripts can be found at [https://github.com/naheld/Held2020\\_TrichoDiel](https://github.com/naheld/Held2020_TrichoDiel).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The mass spectrometry proteomics data has been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifiers PXD016332 and 10.6019/PXD016332 (laboratory experiments) and identifier PXD027796 and 10.6019/PXD027796 (field experiments). The processed proteomic data are also available at the Biological and Chemical Oceanography Data Management Office (BCO-DMO) (<https://www.bco-dmo.org/dataset/783873>). Deployment data for the AT39-05 field expedition can be found at <https://www.bco-dmo.org/deployment/765978>. Source data are provided for main text Figures 1-4 and Extended Data Figures 1-10.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	The diel proteome of cultured <i>Trichodesmium erythraeum</i> sp. IMS101 was investigated via shotgun LC-MS analysis and absolute quantitation of the nitrogenase proteins via parallel reaction monitoring. Triplicate cultures were sampled every 1-3 hours throughout a 24 hour day (n=45) for proteomics, biomass POC and PON content, and nitrogen fixation rates assessed by the acetylene reduction assay.
Research sample	<p>Cultured <i>Trichodesmium erythraeum</i> sp. IMS101, originally isolated from coastal North Carolina and maintained for years in the collection of J. Waterbury. <i>Trichodesmium erythraeum</i> is the most well-studied and genetically characterized <i>Trichodesmium</i> species, is commonly used in laboratory experiments, and represents a common <i>Trichodesmium</i> species found in natural waters. The cultures were not axenic but bacterial contamination was minimized by serial transfer into sterile conditions for many generations prior to the start of the experiment.</p> <p>In addition to the laboratory experiments, field populations of <i>Trichodesmium</i> were sampled from the Eastern subtropical Atlantic ocean at 65 22.240 degrees W, 17 0.284 degrees N over a 24-hour diel cycle on March 10, 2018. Metagenomic analysis suggested that the population was dominated by <i>T. theibautii</i> species.</p>
Sampling strategy	<p>Triplicate experiment vessels containing sterile RMP media were inoculated from the same starting <i>T. erythraeum</i> culture which had prior been maintained in the same conditions as in the experiment. There was no pre-determined statistical calculation of sample size; triplicates were chosen due to incubator space restrictions. The cultures were provided with gentle mixing, oxygenation by sterile air, and a 14:10 light/dark cycle including ramped dawn and dusk transitions at 27°C. The cultures were grown for five days to low-moderate density before diel sampling commenced. From each replicate, distinct samples (n = 45) were collected onto 0.2 µm filters for proteomics (50mL), GFF filters for CHN analysis (10mL) and for acetylene reduction rates (20mL, 1 hour incubation period). Care was taken to reduce exposure to light/darkness during the sampling time, depending on the concurrent incubator light conditions. This study additionally reports the result of an identical, separate experiment conducted one year prior.</p> <p>For field samples, colonies of mixed morphology (puffs and tufts) were collected in biological triplicate by gentle hand-picking from a phytoplankton net, rinsed twice in 0.2µm sterile filtered trace metal clean seawater, then gently decanted onto a 0.2µm supor filter. Samples were flash frozen in liquid nitrogen and stored at -80 degrees C until analysis (proteomics and glycogen content). Care was taken to reduce exposure to light/darkness during sampling time, depending on the time of day.</p>
Data collection	<p>For each laboratory diel sampling time point, N.A.H. or R.M.K. recorded with pen and paper the exact time at the start and end of the sampling procedure and at the end of the acetylene reduction assay incubation period.</p> <p>For field samples, the time of sampling (start and end of phytoplankton net deployment) was recorded with pen and paper by N.A.H. or E.A.W. and additionally in the ship log book. Hand-picking sampling times were recorded by N.A.H. or M.R.M. with pen and paper.</p>
Timing and spatial scale	<p>For the main laboratory experiment, samples were collected continuously over a 24 hour period beginning at 9am EST on December 10, 2019 (one hour prior to "sunrise" in the incubator). Sampling occurred every 1-3 hours to ensure high temporal resolution of the proteome dataset, with the most concentrated sampling occurring around the dawn and dusk transitions. Note that time of day is always reported relative to midnight (5 hours into the dark period) in the incubator.</p> <p>For field samples, samples were taken from a single location (65 22.240 degrees W, 17 0.284 degrees N) and cover a single 24-hour period with sampling occurring every 2-4 hours, similar to the laboratory experiments.</p>
Data exclusions	A sample processing error resulted in poor LC-MS data quality (few proteins identified) for replicate # 2 at four time points (4:00, 8:00, 12:00, 15:00), and this data is therefore not included in the analysis. Due to randomization of the sample processing workflows this is expected to have minimal effect on the study results.
Reproducibility	The results were replicated three times (main triplicated experiment, singlicate replicate experiment, and field sampling). Reproducibility was ensured by triplicate biological replication in the main experiment and is further supported by 1. general and expected biological consistency in the diel patterns and 2. similar findings in an identical yet separate experiment conducted one year prior to the main experiment and 3. similar findings in diel sampling of a field <i>Trichodesmium</i> population. Downstream analysis can be reproduced by accessing the Jupyter Notebook at <a href="https://github.com/naheld/Held2020_TrichoDiel">https://github.com/naheld/Held2020_TrichoDiel</a> .
Randomization	Randomization was not possible due to the time-dependent sampling efforts, however randomization was considered during analysis for POC and PON content and proteomics sample processing and analysis to reduce systematic analytical effects. Specifically, proteomics samples were handled in one batch for each replicate with randomization of the time points within that replicate.
Blinding	For most analyses blinding was not possible as the samples were associated with a specific sampling time point. However, analyses of

## Blinding

POC and PON including an experimental blank were conducted in a blind manner. The clustering and network analyses identified structural patterns in the proteomic data computationally regardless of sample's other characteristics.

Did the study involve field work?  Yes  No

## Field work, collection and transport

## Field conditions

Field populations of *Trichodesmium* were sampled from the Eastern subtropical Atlantic ocean at 65 22.240 degrees W, 17 0.284 degrees N over a 24-hour diel cycle on March 10, 2018 on the AT39-05/Tricolim expedition (R.V. Atlantis, Chief Scientist D. Hutchins, <https://www.bco-dmo.org/deployment/765978>). The region was characterized by low phosphate concentrations (0.13uM at 100m) and high dissolved iron concentration (2.02 nm at 100m) suggesting coastal and/or atmospheric inputs of iron.

## Location

65 22.240 degrees W, 17 0.284 over the entire 24 hour cycle beginning the early morning of March 10, 2018. *Trichodesmium* colonies were sampled from a 130um net size phytoplankton net which was hand deployed to approximately 20m depth then pulled back to the surface five times.

## Access &amp; import/export

The field samples were collected in International waters.

## Disturbance

There was no disturbance to the natural environment in this sampling effort.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

### Methods

- | n/a                                 | Involvement              | Material/System               |
|-------------------------------------|--------------------------|-------------------------------|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Antibodies                    |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Eukaryotic cell lines         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Animals and other organisms   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Human research participants   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Clinical data                 |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Dual use research of concern  |

- | n/a                                 | Involvement              | Method                 |
|-------------------------------------|--------------------------|------------------------|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | MRI-based neuroimaging |