Evidence for polyploidy in the globally important diazotroph *Trichodesmium*

Elizabeth C. Sargent[1,*], Andrew Hitchcock[1,$], S. Andreas Johansson[1,¥], Rebecca Langlois[2], C. Mark Moore[1], Julie LaRoche[2], Alex J. Poulton[3] and Thomas S. Bibby[1*]

[1] Ocean and Earth Science, University of Southampton, National Oceanography Centre Southampton, Southampton, SO14 3ZH, UK.

[2] Department of Biology, Life Science Centre, Dalhousie University, 1355 Oxford Street, Halifax, NS B3H 4R2, Canada.

[3] Ocean Biogeochemistry and Ecosystems, National Oceanography Centre Southampton, Southampton, SO14 3ZH, UK.

[$]Current affiliation: Department of Molecular Biology and Biotechnology, University of Sheffield, Firth Court, Western Bank, Sheffield, S10 2TN, UK.

[¥] Current affiliation: EMBL Heidelberg, Meyerhofstraße 1, 69117 Heidelberg, Germany.

*Corresponding authors:

Elizabeth Sargent. Email: sargenec@mailbox.sc.edu. Current affiliation: Department of Earth and Ocean Sciences, University of South Carolina, Columbia, South Carolina 29208, USA.

Thomas S. Bibby. Email: tsb@noc.soton.ac.uk. Ocean and Earth Science, University of Southampton, National Oceanography Centre Southampton, Southampton, SO14 3ZH, UK.

Running Title: Evidence for polyploidy in *Trichodesmium* (40 characters)

Keywords: *Trichodesmium*, cyanobacteria, polyploidy, *nifH*, nitrogenase, diazotroph

Abstract

Polyploidy is a well-described trait in some prokaryotic organisms; however, it is unusual in marine microbes from oligotrophic environments, which typically display a tendency towards genome streamlining. The biogeochemically significant diazotrophic cyanobacterium *Trichodesmium* is a potential exception. With a relatively large genome and a comparatively high proportion of non-protein-coding DNA *Trichodesmium* appears to allocate relatively more resources to genetic material than closely related organisms and microbes within the same environment. Through simultaneous analysis of gene abundance and direct cell counts we show for the first time that *Trichodesmium spp*. can also be highly polyploid, containing as many as 100 genome copies per cell in field-collected samples and >600 copies per cell in laboratory cultures. These findings have implications for the widespread use of the abundance of the *nifH* gene (encoding a subunit of the N_2-fixing enzyme nitrogenase) as an approach for quantifying the abundance and distribution of marine diazotrophs. Moreover, polyploidy may combine with the unusual genomic characteristics of this genus both in reflecting evolutionary dynamics and influencing phenotypic plasticity and ecological resilience.

173 words

Introduction

Smaller genomes and reduced gene diversity are characteristics of many microbes adapted to life in the oligotrophic oceans (Swan *et al*., 2013). Such a strategy can be advantageous under nutrient-limited conditions as fewer resources are required to maintain and duplicate a genome, but comes at the cost of reduced physiological flexibility (Yooseph *et al.*, 2010). In stark contrast is the cyanobacterium *Trichodesmium*, a genus of colony-forming marine diazotrophs prolific in N_2-fixation in oligotrophic tropical and sub-tropical oceans (Capone *et al*., 2005). *Trichodesmium* has a large (7.75 Mbp) genome with low protein coding capacity (~40% non-protein-coding DNA) and a high level of gene duplication (10% of all genes) (Bergman *et al*. 2013; Walworth *et al*. 2015). The abundant non-protein-coding sequences, over 80% of which are transcribed, consist of a combination of non-coding RNAs (ncRNAs), selfish DNA elements, transposases and introns, which possibly contribute to the genome expansion and metabolic flexibility observed in this group (Pfreundt *et al*. 2014; 2015; Walworth *et al*., 2015). These features may also contribute to the versatility and ecological success of *Trichodesmium* (Bergman *et al*. 2013; Pfreundt *et al*. 2014; Walworth *et al*., 2015).

In contrast to the small-celled picocyanobacterial lineages *Prochlorococcus* and *Synechococcus*, which co-inhabit and numerically dominate low latitude oligotrophic marine environments, cell sizes are also much greater for species of *Trichodesmium*, being >3 orders of magnitude larger by volume (Carpenter *et al*. 2004). Moreover, *Trichodesmium* are often found in association with a diverse community of other microbes (Pearl et al. 1989; Hewson et al. 2009), factors indicating that *Trichodesmium* is an unusual and unique inhabitant in oligotrophic tropical environments (Walworth *et al.*, 2015).

Polyploidy, the presence of multiple genome copies per cell, has received little assessment in marine microbes. Maintaining additional genome copies represents a significant nutrient investment of both nitrogen and phosphorus, which is a scarce resource in the oligotrophic ocean (Elser *et al.*, 2003; Karl, 2014), with the latter nutrient potentially particularly important in constraining the growth of diazotrophs such as *Trichodesmium* (Sanudo-Wilhelmy *et al*. 2001). Despite this, polyploidy in cyanobacteria has been widely documented with multiple genome copies per cell commonly reported for a variety of model and ecologically important species (Griese *et al*. 2011, Zerulla *et al*. 2016). In this study we investigated polyploidy in *Trichodesmium* and the implications this may have for both the success of this genera and the use of gene copy numbers in assessing the biogeography and abundance of marine diazotrophic species.

Materials and methods

Sample collection

Culture: Trichodesmium IMS101 was grown in YBC-II medium under a 12/12-h light/dark cycle at 25°C (Richier *et al.*, 2012). For DNA analysis 10 ml samples from triplicate exponential phase cultures were filtered onto 0.22-µm Durapore (Millipore) filters under low (2 mbar) vacuum pressure. After filtration, filters were flash frozen in liquid nitrogen and stored at -80˚C until DNA extraction. For cell counts, 10 mL of culture was collected in parallel from triplicate exponential phase cultures and preserved in 2% acidic Lugol's iodine (Throndsen, 1978).

Field: Samples were collected during the AMT17 (Oct-Nov 2005) and D361 (Feb-Mar 2011) research cruises in the tropical and subtropical Atlantic (see Snow *et al*. 2015 for

cruise tracks). During these cruises, *Trichodesmium* cell count samples were collected through filtration of a 20 L surface CTD bottle drained through a 10 µm polycarbonate filter. The resulting retentate was rinsed into a 50 mL amber bottle using 0.2 µm filtered seawater, and was preserved with 2% acidic Lugol's iodine (Throndsen, 1978). For DNA analysis, 2 L of seawater was collected from a replicate surface CTD bottle and was filtered in duplicate onto 0.22 µm Durapore (Millipore) filters under low (2 mbar) vacuum pressure. After filtration, filters were flash frozen in liquid nitrogen and stored at -80°C until DNA extraction.

DNA extraction, gene abundance and genome copy

Frozen filters were crushed using a nucleic acid free, sterile plastic pestle and DNA was extracted using the Qiagen DNeasy mini plant kit according to the manufacturer's protocol. Following extraction, DNA concentrations were determined using the RediPlate 96 dsDNA Quantitation Kit (Molecular Probes) and read on a Fluoroscan Ascent microplate reader.

Quantitative (q)PCR was performed using primers and a probe chosen for amplification of the filamentous *nifH* phylotypes (Langlois *et al.*, 2008), or primers designed to specifically amplify predicted double or triple copy genes from the *Trichodesmium erythraeum* IMS101 genome (Table S1). For the multi-copy number genes the primers were chosen such that they would amplify the same product from each of the individual copies. For *nifH* analysis, qPCRs were run on an ABI Prism 7000 (Applied Biosystems) using cycling conditions and reaction parameters as described previously (Langlois *et al*. 2008). For analysing the ratio of *nifH* to other genes, qPCR was performed using a Mx3005P qPCR System using Brilliant III Ulta-FAST SYBR Green QPCR Master Mix (Agilent Technologies,

Santa Clara CA, USA) with a thermal profile of an initial denaturation at 95°C for 3 min followed by 40 cycles of denaturation for 15 s at 95°C and combined annealing/elongation for 20 s at 60°C. Dissociation curves were collected between 55-95°C and revealed a single melting point, which was confirmed to be due to amplification of single products by agarose gel electrophoresis. The gene copy number ratio was estimated using the mean (n=3) difference in Ct-value from a gDNA standard curve (2-fold serial dilution from 1 to 1:64) for each primer pair and the primer efficiency calculated over the same dilution series. Standard curves all had R^2 values for linearity > 0.99 and primer efficiencies between 102-106%. The formula for calculating the ratio between two given genes A and B was:

(Primer efficiency gene A^CT gene A)/(Primer efficiency gene B^CT gene B).

Cell counts, chlorophyll concentration and DNA staining

Cell abundances for *Trichodesmium* were directly measured from free trichomes via light microscopy at 200x magnification. Chlorophyll concentrations were obtained from culture studies using 10 ml of cell cultures filtered onto Whatman GF/F filters, which were flash frozen in liquid nitrogen and stored at -20°C until extraction and fluorometric chlorophyll determination (Welschmeyer, 1994). For field sampling, *in situ* community chlorophyll concentrations were measured through collecting 200 mL of seawater from a surface CTD bottle, which was filtered onto a Whatman GF/F filter and then extracted immediately. Extraction was performed in the dark in 8 mL 90% acetone overnight at 4˚C, with subsequent concentrations in the extract determined using a TD-700 fluorometer calibrated using an RS Aqua red solid standard.

Trichodesmium erythraeum IMS101 was examined under confocal microscopy following 4',6-diamidino-2-phenylindole (DAPI) staining to assess intracellular DNA distribution. Following gravitational settling and aspiration of 5 mL culture, *Trichodesmium* biomass was fixed with 1% glutaraldehyde, 3% formaldehyde, 14% sucrose in piperazine-N,N′-bis(2-ethanesulfonic acid) (PIPES) buffer. After 20 minutes at 4°C, samples were filtered onto 0.8 µm polycarbonate black filters, transferred to microscope slides, and mounted with VECTASHIELD HardSet Mounting Medium with DAPI (Vector Laboratories) at 1.5 µg mL^{-1}. Slides were allowed to set overnight at 4°C. Samples were imaged on a Leica SP5 confocal microscope under UV excitation, which excites DAPI associated with DNA, and 488 nm excitation, which excites chlorophyll *a* and phycoerythrin. DAPI associated with RNA is also excited at this wavelength, but the associated emission spectrum is weak (Suzuki *et al.* 1997).

Results and discussion

Comparison of gene and cellular abundance

Direct measurement of *nif*H gene copies (*nif*H L^{-1}; as described in Langlois *et al*., 2008) and cell counts (cells L^{-1}) of *in situ Trichodesmium* populations, sampled from two research cruises in the Atlantic Ocean and laboratory-grown cultures, revealed that *nifH* abundance exceeds cell abundance by 1-2 orders of magnitude (Table 1 and Figure 1). Thus, despite a strong correlation between cell count based abundances and *nif*H abundances (R^2 = 0.89) (Figure 1) the latter considerably exceed the former, which is consistent with unexplained discrepancies previously reported in the literature (Luo *et al.*, 2012; Rouco *et al*. 2014).

As *nifH* is a single copy gene in the *Trichodesmium* genome (Zehr *et al*., 2008; Walworth *et al*. 2015), we estimated the degree of ploidy (genome copies $cell^{-1}$) using an established approach (genome copies $cell^{-1}$ = *nifH* L^{-1} / cells L^{-1}; Pecoraro *et al*. 2011). The estimated degree of ploidy ranged from 1-120 genome copies per cell for *in situ* samples (n=31) excluding those samples where <1 copy number per cell was estimated. Errors in both cell counts and copy number estimates may have been more significant for such samples, which were all at low overall biomass (Figure 1). For example, the low total copy numbers may have been too low for the quantitative PCR (qPCR) technique to work reliably. Additionally, dead cells with partially degraded DNA may also have contributed disproportionately within the low biomass field sampled locations. Estimated copies per cell exceeded 600 in laboratory-cultured *Trichodesmium erythraeum* IMS 101 (n=2) (Table 1). Thus, both field and cultured *Trichodesmium* displayed substantial, albeit variable, degrees of polyploidy, with our limited dataset further suggesting that *nifH* copy number per cell was higher in culture as opposed to field collected samples.

To confirm that the observed mismatch between measured gene abundances and cell counts was not specific to the *nifH* amplicon, genes present in the *Trichodesmium* IMS101 genome in double (16S rRNA) or triple (*psb*A) copies were also analyzed by qPCR from cultured samples. Our measured ratios (Table 2) compared well with the predicted copy numbers from the genome, arguing against cryptic amplification of *nifH* copies and suggesting robust estimates of the degree of ploidy.

Localisation of DNA in Trichodesmium cells

To visualise the localisation of DNA in *Trichodesmium* cells, confocal microscopy following DAPI staining of *T. erythraeum* IMS101 DNA was performed. The markedly segregated DAPI staining throughout the cell implies intracellular DNA distribution in *T. erythraeum* IMS101 is scattered and extensive (Figures 1B and 1C); an observation that is consistent with *Trichodesmium* possessing multiple copies of the genome per cell, as have been reported in other cyanobacteia (Schneider *et al.*, 2007; Lane and Martin 2010; Sukenik *et al.* 2012; Zerulla et al. 2016).

Implications of polyploidy for abundance estimates

Quantitative (q)PCR is commonly used to quantify the environmental abundance of the highly conserved *nif*H gene that encodes the iron-binding component of the nitrogenase enzyme present in all diazotrophs (Zehr *et al.* 1998; Luo *et al*. 2012). Such approaches can be used to assess the biogeography of diazotrophic phylotypes and have been crucial in increasing our understanding of the intricacies of the marine nitrogen cycle (Fernández *et al.*, 2010; Goebel *et al.*, 2010; Zehr, 2011; Thompson and Zehr, 2013). In the Atlantic Ocean, *Trichodesmium nif*H constitutes as much as 50% of the total detectable *nif*H (Langlois *et al.*, 2008) and can exceed contributions by unicellular diazotrophic phylotypes by up to an order of magnitude (Rijkenberg *et al.*, 2011). However, in the case of *Trichodesmium*, failure to account for any degree of polyploidy (Table 1) would result in an overestimate of cell abundance if extrapolations were made from gene abundance to cell abundance assuming a 1:1 cell:genome ratio.

To further highlight this potential issue and provide additional circumstantial evidence for significant polyploidy within field communities, we extrapolated estimates of

199 *Trichodesmium* abundances to corresponding *in situ* chlorophyll concentrations (Table 3).
200 For example, using a typical cellular chlorophyll content of 1.1 pg Chl per cell (LaRoche and
201 Breitbarth, 2005) we can estimate the contribution of *Trichodesimum* to total measured *in*
202 *situ* whole community chlorophyll from a range of published data alongside that collected
203 within the current study (Table 3). Consistent with previous estimates (Carpenter *et al*.
204 2004), *Trichodesmium* accounted for as much as 30% of measured *in situ* whole community
205 chlorophyll when estimated from cell counts. In contrast, performing a similar calculation
206 based on measured *nifH* gene copies and assuming only one *nifH* copy per cell (i.e. no
207 polyploidy) would result in a ten-fold increase in the average *Trichodesmium* contribution to
208 total chlorophyll (Table 3). Moreover, under an assumption of monoploidy, *Trichodesmium*
209 alone might be estimated to contribute up to an order of magnitude higher chlorophyll than
210 the actual measured community chlorophyll concentration, highlighting that the number of
211 *nifH* copies per cell must be >1.
212
213 ***Ecophysiological implications of polyploidy in Trichodesmium spp.***
214 There may be several ecological advantages to polyploidy in *Trichodesmium* that would
215 overcome the obvious disadvantages of maintaining multiple genome copies per cell. For
216 instance, multiple genome copies distributed throughout each cell (suggested in Figure 1B
217 and C) would allow efficient transcription in localised areas of the cell. This is an
218 advantageous feature, as *Trichodesmium* needs to simultaneously reconcile oxygen-evolving
219 photosynthesis and anaerobic nitrogen fixation processes within a large cellular volume
220 (Bergman *et al.*, 2013). Being polyploid is thought to enable quick instigation of metabolic
221 activity following dormancy and also long-term preservation of genome integrity in

222 *Aphanizomenon ovalisporum* (Sukenik *et al.*, 2012). The ability to quickly instigate
223 alterations in metabolism through polyploidy may be a considerable advantage for
224 *Trichodesmium,* enabling exploitation of ephemeral nutrient pulses, such as through aeolian
225 iron deposition which is a key factor in the biogeography of this group (Moore *et al*., 2009).
226 While there are considerable advantages to polyploidy, there are also costs associated
227 with the higher resource requirement to maintain and duplicate multiple genomes. The
228 considerable range of polyploidy observed in *Trichodesmium*, may reflect variability with
229 growth phase, diel cycle, nutrient availability or intercellular variability within a colony (e.g.
230 the presence of diazocytes). The increased nitrogen requirement for multiple genome copies
231 is presumably not as disadvantageous for *Trichodesmium* as it would be for non-diazotrophic
232 organisms such as *Prochlorococcus* and *Synechococcus*. However, considering the potential
233 range in genome copies per cell (Table 1), using an estimated genomic phosphorus content
234 of 8 x 10^{-4} pg genome^{-1} (Walworth *et al.*, 2015) and a typical range of intracellular
235 phosphorus of 0.5-1 pg cell^{-1}, (Nuester *et al*. 2012; Tovar-Sanchez & Sañudo-Wilhelmy 2011),
236 100 genome copies per cell (maximum estimated in marine environment, Table 1), would
237 account for 8-16% of the total cellular phosphorus content. The degree of polyploidy may
238 therefore significantly impact the phosphorus resource allocation in *Trichodesmium*
239 (Sanudo-Wilhelmy *et al*. 2001; Elser *et al.*, 2003; Nuester *et al.*, 2012). Indeed recent results
240 on the single-celled cyanobacteria *Synechocystis* sp. suggest that phosphorus availability has
241 an impact on the degree of ploidy observed in this cyanobacterial species (Zerulla *et al.*
242 2016). Our data reveal higher polyploidy within phosphorous-replete cultures compared to
243 field samples collected within low phosphorous environments (Sanudo-Wilhelmy *et al*.

2001; Moore *et al*. 2009), which is consistent with such a response (Figure 1), although further work would clearly be required to substantiate such suggestions.

In summary, we provide direct evidence linking gene abundance to cell number that indicates that *Trichodesmium* is highly polyploid in culture and exhibits a highly diverse degree of ploidy in the field. This observation highlights that caution should be applied in extrapolating *nifH* gene abundance data to estimate diazotrophic cell abundances and suggests that a wider assessment of the extent of polyploidy in ecologically significant marine (cyano-)bacterial taxa would be desirable. A high degree of polyploidy in *Trichdesmium* adds to the unique genomic characteristics of this organism when compared to the majority of cyanobacteria in oligotrophic oceanic environments (Walworth *et al.*, 2015). *Trichodesmium* appears to have evolved to capitalise on some of the benefits of maintaining multiple copies of a large genome; features that may enable *Trichodesmium* to continue to be a prolific and robust player in a changing future ocean. Further work is required to determine how environmental conditions, growth phase and cell differentiation may impact the degree and function of ploidy observed in this keystone oceanic microbe.

Acknowledgements

This study is a contribution to the international IMBER project and was supported by the UK Natural Environment Research Council National Capability funding to Plymouth Marine Laboratory and the National Oceanography Centre. This is contribution number ### of the AMT programme. We thank the captain and crew of the RRS James Cook and the RRS Discovery. This work was supported by a UoS GSNOCS PhD Scholarship awarded to ES, Natural Environmental Research Council National Capability funding allocated to AP, and DFG (R02138/5-1) and BMBF (SOPRAN) grants to JLR. We also thank Joseph Snow and Claire Mahaffey for useful discussions relating to the development of the current work.

Conflict of interest statement

The authors declare no conflict of interest.

References

Altshul SF, Madden TL, Schaffer AA, Zhang J, Miller W *et al*. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**(1):3389-402.

Bergman B, Sandh G, Lin S, Larsson J, Carpenter EJ. (2013). Trichodesmium--a widespread marine cyanobacterium with unusual nitrogen fixation properties. *FEMS Microbiol Rev* **37**:286–302.

284 Capone DG, Burns JA., Montoya JP, Subramaniam A, Mahaffey C, Gunderson T, *et al.* (2005).
285 Nitrogen fixation by Trichodesmium spp.: An important source of new nitrogen to the
286 tropical and subtropical North Atlantic Ocean. *Global Biogeochem Cycles* **19**: GB2024.

287 Carpenter EJ, Subramaniam A, Capone DG. (2004). Biomass and primary productivity of the
288 cyanobacterium Trichodesmium spp. in the tropical N Atlantic ocean. *Deep-Sea Res Pt I* **51**;
289 173–203.

290 Elser JJ, Acharya K, Kyle M, Cotner J, Makino W, Markow T, *et al.* (2003). Growth rate-
291 stoichiometry couplings in diverse biota. *Ecol Lett* **6**:936–943.

292 Fernández a., Mouriño-Carballido B, Bode a., Varela M, Marañón E. (2010). Latitudinal
293 distribution of *Trichodesmium* spp. and N_2 fixation in the Atlantic Ocean. *Biogeosciences*
294 **7**:3167–3176.

295 Goebel NL, Turk K a, Achilles KM, Paerl R, Hewson I, Morrison AE, *et al.* (2010). Abundance
296 and distribution of major groups of diazotrophic cyanobacteria and their potential
297 contribution to N_2 fixation in the tropical Atlantic Ocean. *Environ Microbiol* **12**:3272–89.

298 Griese M, Lange C, Soppa J. (2011). Ploidy in cyanobacteria. *FEMS Microbiol Lett* **323**:124–
299 31.

300 Karl DM. (2014). Microbially mediated transformations of phosphorus in the sea: new views
301 of an old cycle. *Ann Rev Mar Sci* **6**:279–337.

302 Lane N, Martin W. (2010). The energetics of genome complexity. *Nature Hypothesis* **467**:929-
303 934.

304 Langlois RJ, Hümmer D, LaRoche J. (2008). Abundances and distributions of the dominant
305 nifH phylotypes in the Northern Atlantic Ocean. *Appl Environ Microbiol* **74**:1922–31.

306 LaRoche J, Breitbarth E. (2005). Importance of the diazotrophs as a source of new nitrogen
307 in the ocean. *J Sea Res* **53**:67–91.

308 Larsson J, Nylander JA, Bergman B. (2011). Genome fluctuations in cyanobacteria reflect
309 evolutionary, developmental and adaptive traits. *BMC Evol Biol* **11**:187–209.

310 Luo Y-W, Doney SC, Anderson L a., Benavides M, Berman-Frank I, Bode A, *et al.* (2012).
311 Database of diazotrophs in global ocean: abundance, biomass and nitrogen fixation rates.
312 *Earth Syst Sci Data* **4**:47–73.

313 Moore CM, Mills MM, Achterberg EP, Geider RJ, LaRoche J, Lucas MI, *et al.* (2009). Large-scale
314 distribution of Atlantic nitrogen fixation controlled by iron availability. *Nat Geosci* **2**:867–
315 871.

316 Nuester J, Vogt S, Newville M, Kustka AB, Twining BS. (2012). The unique biogeochemical
317 signature of the marine diazotroph trichodesmium. *Front Microbiol* **3**:150.

318 Pecoraro V, Zerulla K, Lange C, Soppa J. (2011). Quantification of ploidy in proteobacteria
319 revealed the existence of monoploid, (mero-)oligoploid and polyploid species. *PLoS One*
320 **6**:e16392.

321 Pfreundt U, Kopf M, Belkin N, Berman-Frank I, Hess WR. (2014). The primary transcriptome
322 of the marine diazotroph Trichodesmium erythraeum IMS101. *Sci Rep* **4**:6187.

Pfreundt U, Hess WR. (2015). Sequential splicing of a group II twintron in the marine cyanobacteriumTrichodesmium. *Sci Rep* **5**:16829.

Richier S, Macey AI, Pratt NJ, Honey DJ, Moore CM, Bibby TS. (2012). Abundances of iron-binding photosynthetic and nitrogen-fixing proteins of Trichodesmium both in culture and in situ from the North Atlantic. *PLoS One* **7**:e35571.

Rijkenberg MJ a, Langlois RJ, Mills MM, Patey MD, Hill PG, Nielsdóttir MC, *et al.* (2011). Environmental forcing of nitrogen fixation in the eastern tropical and sub-tropical North Atlantic Ocean. *PLoS One* **6**:e28989.

Rouco M, Warren HJ, McGillicuddy DJ, Waterbury JB, Dyhrman ST. (2014). Trichodesmium sp. clade distributions in the western North Atlantic Ocean. *Limnol Oceanogr* **59**:1899–1909.

Sañudo-Wilhelmy SA, Kustka AB, Gobler CJ, Hutchins DA, Yang M, Lwiza K *et al.* (2001). Phosphorus limitation of nitrogen fixation by Trichodesmium in the central Atlantic Ocean. *Nature* **411**: 66-69.

Schneider D, Fuhrmann E, Scholz I, Hess WR, Graumann PL. (2007). Fluorescence staining of live cyanobacterial cells suggest non-stringent chromosome segregation and absence of a connection between cytoplasmic and thylakoid membranes. *BMC Cell Biol* **8**:39.

Snow JT, Schlosser C, Woodward EMS, Mills MM, Achterberg EP, *et al*. (2015). Environmental controls on the biogeography of diazotrophy and *Trichodesmium* in the Atlantic Ocean. *Global Biogeochem* **29**:865-884.

342 Sukenik A, Kaplan-Levy RN, Welch JM, Post AF. (2012). Massive multiplication of genome
343 and ribosomes in dormant cells (akinetes) of Aphanizomenon ovalisporum (Cyanobacteria).
344 *ISME J* **6**:670–9.

345 Suzuki T, Fujikura K, Higashiyama T and Takata K (1997). DNA staining of fluorescence and
346 laser confocal microscopy. *J Histochem & Cytochem* **45**(1):49-53

347 Swan BK, Tupper B, Sczyrba A, Lauro FM, Martinez-garcia M, González JM, *et al.* (2013).
348 Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the
349 surface ocean. *Proc Natl Acad Sci U S A* **110**:11463–8.

350 Thompson AW, Zehr JP. (2013). Cellular interactions: lessons from the nitrogen-fixing
351 cyanobacteria. *J Phycol* **49**:1024–1035.

352 Tovar-Sanchez A, Sañudo-Wilhelmy SA. (2011). Influence of the Amazon River on dissolved
353 and intra-cellular metal concentrations in *Trichodesmium* colonies along the western
354 boundary of the sub-tropical North Atlantic Ocean. *Biogeosciences* **8**:217–225.

355 Throndsen J (1978). P and storage. P manual. AS 69-74. (1978). Preservation and storage.
356 Phytoplankton manual. A. Sournia. A Sournia 69–74.

357 Walworth N, Pfreundt U, Nelson WC, Mincer T, Heidelberg JF, Fu F *et al.* (2015).
358 Trichodesmium genome maintains abundant, widespread noncoding DNA in situ, despite
359 oligotrophic lifestyle. *Proc Natl Acad Sci USA* **112**: 4251-56.

360 Welschmeyer N. (1994). Fluorometric analysis of chlorophyll a in the presence of chlorophyll
361 b and pheopigments. *Limnol Oceanogr* **39**:1985–1992.

362 Yooseph S, Nealson KH, Rusch DB, McCrow JP, Dupont CL, Kim M, *et al.* (2010). Genomic and
363 functional adaptation in surface ocean planktonic prokaryotes. *Nature* **468**:60–6.

364 Zehr JP, Mellon MT, and Zani S. (1998). New nitrogen-fixing microorganisms detected in
365 oligotrohpic oceans by amplification of nitrogenase (*nif*H) genes. *App Env Microbio*
366 **64**(9):3444-3450.

367 Zehr JP, Bench SR, Carter BJ, Hewson I, Niazi F, *et al.* (2008). Globally distributed
368 uncultivated oceanic N2-fixing cyanobacteria lack oxygenic photosystem II. Science 322:
369 1110-1112 doi:10.1126/science.1165340, 2008

370 Zehr JP. (2011). Nitrogen fixation by marine cyanobacteria. *Trends Microbiol* **19**:162–73.

371 Zerulla K, Ludt K, Soppa J (2016). The ploidy level of Synechocystis sp. PCC 6803 is highly
372 variable and is influenced by growth phase and by chemical and physical external parameters.
373 Microbiology. 162:730-739 doi: 10.1099/mic.0.000264. Epub 2016 Feb 25.

374

375 **Table and figure legends**

376

377 **Table 1. Estimated genome copies in *Trichodesmium* spp.** Genome copies per cell are
378 calculated by division of gene copies L^{-1} by cells L^{-1}, as described by Pecoraro *et al*. (2011).
379 The table shows the ranges of cell abundance and *nifH* gene copy numbers measured from
380 cultures of *Trichodesmium* or surface seawater samples collected from two research cruises
381 to the Atlantic (AMT17 and D361). The range of derived genome copies per cell (i.e. level of
382 ploidy) is also shown. Sampling procedures for cultures and *in situ* samples, as well as
383 methods for measurement of cell count by microscopy and gene abundance by qPCR, are
384 described in material and methods.

385

386 **Table 2. Predicted and measured ratio of multi-copy number genes in *Trichodesmium***
387 ***erythraeum* ISM101 genome.** The *Trichodesmium erythraeum* ISM101 genome predicts
388 that the *nifH*, 16S rRNA and *psbAII* genes are present in the *T. erythraeum* ISM101genome
389 (Walworth *et al.* 2015) in single (Tery_4136), double (Tery_R0014, Tery_R0029) and triple
390 (Tery_0182, Tery_0183, Tery_4763) copies respectively. The abundances of these genes in
391 cultures of *T. erythraeum* ISM101 were measured using qPCR from DNA extracted from the
392 same culture. The corresponding measured ratio of these genes is shown as an average and
393 standard deviation of n=3 samples.

394

395 **Table 3. Calculated potential contributions to total *in situ* chlorophyll by**
396 ***Trichodesmium spp***. The calculated contribution that *Trichodesmium* makes to measured *in*
397 *situ* whole community chlorophyll derived from cells counts and gene copies, with the latter

assuming monoploidy (i.e. assuming 1 copy per cell), from published data and this study. Conservative estimates are made using the lowest reported chlorophyll content per cell (1.1. pg Chl per cell, La Roche and Breitbarth 2005). Note that the large variation in gene copy-based contributions to total chlorophyll under the assumption of monoploidy are likely due to variability in the actual level of ploidy in natural populations of *Trichodesmium*.

Figure 1. Relationship between *Trichodesmium* cell counts and *nifH* gene copies (Table 1). (**A**) A combined assessment of *in situ* data from surface samples along the AMT17 transect (blue) and D361 transect (black) and *Trichodesmium* erythraeum IMS101 culture samples (red). A significant correlation is observed ($R^2 = 0.89$, $y = 0.5423x + 0.9624$) Gene copies consistently exceed cell counts by 1-2 orders of magnitude and thus deviated from a 1:1 correlation (dotted line). (**B**) Confocal microscopy of DAPI stained *T. erythraeum* IMS101 cultures observed under UV and 488 nm excitation. DNA (blue) and chlorophyll autofluorescence (orange) in a single *T. erythraeum* IMS101 cell demonstrating marked scattered intracellular DNA distribution. (**C**) Scattered DNA distribution was observed in cells along a trichome. Scale bars represent 2 μm, variation in intensity is due to changes in cellular morphology along the z-plane. See materials and methods for detailed microscopy methods.

417 **Tables**

418

419 **Table 1**

***Trichodesmium* Sample**	***nifH* copies L^{-1}**	**cells L^{-1}**	**Genome copies per cell [range (avg ± SD)]**
Cruise AMT17 (n=15)	ND – 2.7E+5	56 – 1.8E+4	1-120 (12 ± 13)
Cruise D361 (n=16)	ND – 6.7E+5	0 – 1.9E+4	1-50 (31 ± 30)
Trichodesmium erythraeum IMS 101 (n=2)	2.5E+7 – 1.4E+8	3.6E+4 – 2.1E+5	639-697 (668 ± 41)

420

421

422

423 **Table 2**

Gene targets	**Predicted ratio**	**Measured ratio**
nifH:16S rRNA	1:2	1:1.53 (±0.06)
nifH:*psbAII*	1:3	1:3.32 (±0.13)

424

425 **Table 3**

***Trichodesmium* contribution to total chlorophyll (%) estimated from either cell counts or gene copies (assuming monoploidy)**					
Cell Counts		**Gene Copies**			
Range	**Avg ± SD**	**Range**	**Avg ± SD**	**n**	**Source**
0-8.0	4.1 ± 1.7	NA	NA	16	This study, AMT 21
0-11.6	4.4 ± 7.9	NA	NA	33	Fernández et al. 2010**
0.2-13.7	7.2 ± 4.5	NA	NA	22	Letelier & Karl 1996**
0-175.5	18.8 ± 30.3	NA	NA	336	Capone et al. 2004**
0-243.9	22.9 ± 34.0	NA	NA	335	Borstad 1978**
0-6	2.1 ± 2.5	0-98.4	10 ± 24.9	31	This study, AMT 17
0-17.8	4.1 ± 4.3	0-138.7	44.3 ± 63.9	15	This study, D361
NA	NA	1-440	136.0 ± 126.1	13	Goebel et al. 2010*
NA	NA	0-1163.8	99.6 ± 275.6	23	Kong et al. 2011**
NA	NA	0-2065.5	120.3 ± 319.1	125	Moisander et al. 2010**

*Data estimated from figures, **Data from Luo *et al.* (2012) dataset.

426
427

Figure 1

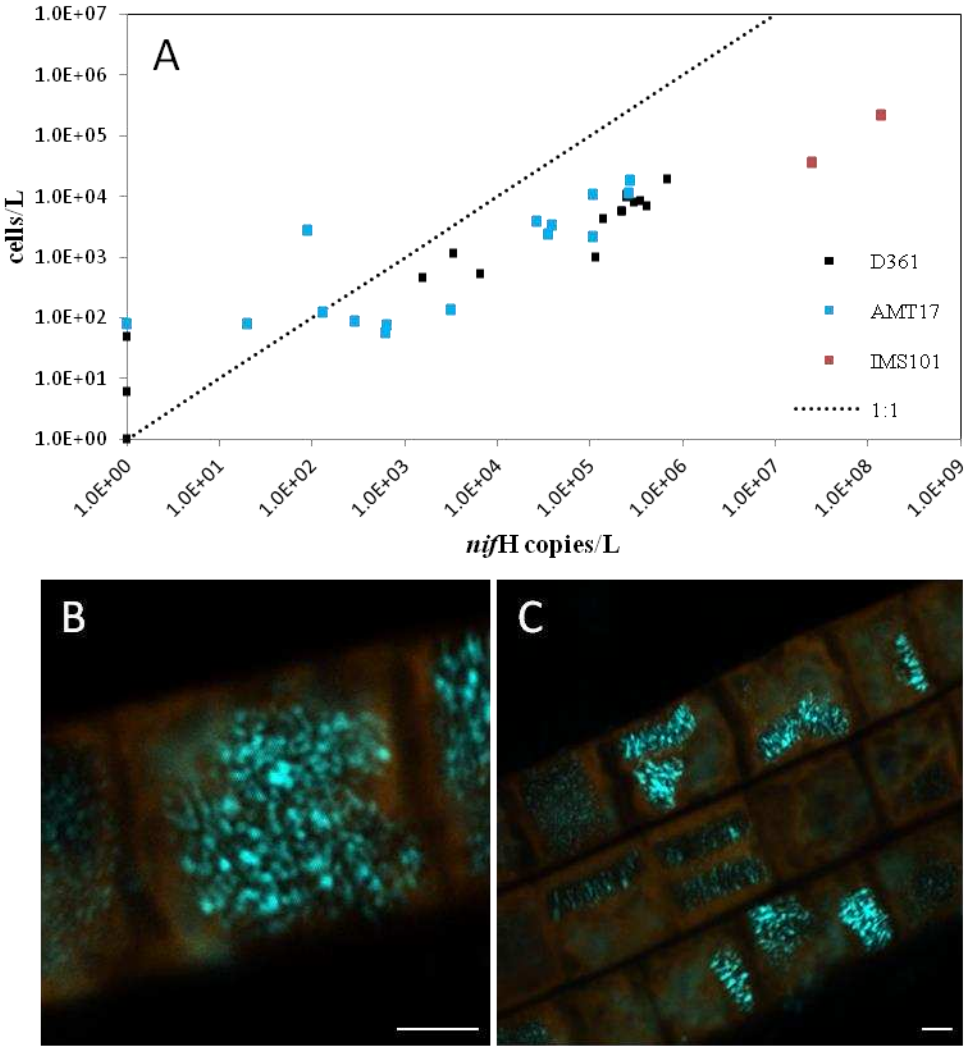
A
cells/L
1.0E+07
1.0E+06
1.0E+05
1.0E+04
1.0E+03
1.0E+02
1.0E+01
1.0E+00
1.0E+00
1.0E+01
1.0E+02
1.0E+03
1.0E+04
1.0E+05
1.0E+06
1.0E+07
1.0E+08
1.0E+09
nifH copies/L
D361
AMT17
IMS101
1:1
B
C

Table S1. Genes analysed by quantitative PCR in this study and corresponding *Trichodesmium*-specific qPCR oligonucleotide primers. Primer specificity determined by BLAST analysis (Altschul *et al*. 1997).

Gene	Homologues in *T. erythraeum* IMS 101[a]	Primer	Sequence (5′-3′)	Amplicon size (bp)
*nif*H	*Tery_4136*	Forward	TGGCCGTGGTATTATTACTGCTATC	111
		Reverse	GCAAATCCACCGCAAACAAC	
		Probe	AAGGAGCTTATACAGATCTA	-
16S rRNA	*Tery_R0014, Tery_R0029*	Forward	CCACACTGGGACTGAGACAC	117
		Reverse	AACCCTAGAGCCTTCCTCCC	
*psb*A	*Tery_0182, Tery_0183, Tery_4763*	Forward	AATGCACCCATTCCACATGC	192
		Reverse	CGACCGAAGTAGCCATGAG	

[a] annotated in the *Trichodesmium erythraeum* IMS 101 genome (Genbank accession NC_008312).

434