Evidence for polyploidy in the globally important diazotroph *Trichodesmium*

Elizabeth C. Sargent¹*, Andrew Hitchcock¹, S. Andreas Johansson¹, Rebecca Langlois², C. Mark Moore¹, Julie LaRoche², Alex J. Poulton³ and Thomas S. Bibby¹*

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 N₂-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:

$\frac{(\text{Primer efficiency gene A}^\text{CT gene A})}{(\text{Primer efficiency gene B}^\text{CT gene B})}$.

**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 *nifH* gene copies (nifH 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 *nifH* 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 (*psbA*) 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 cyanobacteria (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 *nifH* 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 nifH* constitutes as much as 50% of the total detectable *nifH* (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
Trichodesmium abundances to corresponding *in situ* chlorophyll concentrations (Table 3). For example, using a typical cellular chlorophyll content of 1.1 pg Chl per cell (LaRoche and Breitbarth, 2005) we can estimate the contribution of *Trichodesmium* to total measured *in situ* whole community chlorophyll from a range of published data alongside that collected within the current study (Table 3). Consistent with previous estimates (Carpenter *et al.*, 2004), *Trichodesmium* accounted for as much as 30% of measured *in situ* whole community chlorophyll when estimated from cell counts. In contrast, performing a similar calculation based on measured *nifH* gene copies and assuming only one *nifH* copy per cell (i.e. no polyploidy) would result in a ten-fold increase in the average *Trichodesmium* contribution to total chlorophyll (Table 3). Moreover, under an assumption of monoploidy, *Trichodesmium* alone might be estimated to contribute up to an order of magnitude higher chlorophyll than the actual measured community chlorophyll concentration, highlighting that the number of *nifH* copies per cell must be $>1$.

**Ecophysiological implications of polyploidy in Trichodesmium spp.**

There may be several ecological advantages to polyploidy in *Trichodesmium* that would overcome the obvious disadvantages of maintaining multiple genome copies per cell. For instance, multiple genome copies distributed throughout each cell (suggested in Figure 1B and C) would allow efficient transcription in localised areas of the cell. This is an advantageous feature, as *Trichodesmium* needs to simultaneously reconcile oxygen-evolving photosynthesis and anaerobic nitrogen fixation processes within a large cellular volume (Bergman *et al.*, 2013). Being polyploid is thought to enable quick instigation of metabolic activity following dormancy and also long-term preservation of genome integrity in
Aphanizomenon ovalisporum (Sukenik et al., 2012). The ability to quickly instigate alterations in metabolism through polyploidy may be a considerable advantage for Trichodesmium, enabling exploitation of ephemeral nutrient pulses, such as through aeolian iron deposition which is a key factor in the biogeography of this group (Moore et al., 2009).

While there are considerable advantages to polyploidy, there are also costs associated with the higher resource requirement to maintain and duplicate multiple genomes. The considerable range of polyploidy observed in Trichodesmium, may reflect variability with growth phase, diel cycle, nutrient availability or intercellular variability within a colony (e.g. the presence of diazocytes). The increased nitrogen requirement for multiple genome copies is presumably not as disadvantageous for Trichodesmium as it would be for non-diazotrophic organisms such as Prochlorococcus and Synechococcus. However, considering the potential range in genome copies per cell (Table 1), using an estimated genomic phosphorus content of 8 x 10^-4 pg genome^-1 (Walworth et al., 2015) and a typical range of intracellular phosphorus of 0.5-1 pg cell^-1, (Nuester et al. 2012; Tovar-Sanchez & Sañudo-Wilhelmy 2011), 100 genome copies per cell (maximum estimated in marine environment, Table 1), would account for 8-16% of the total cellular phosphorus content. The degree of polyploidy may therefore significantly impact the phosphorus resource allocation in Trichodesmium (Sanudo-Wilhelmy et al. 2001; Elser et al., 2003; Nuester et al., 2012). Indeed recent results on the single-celled cyanobacteria Synechocystis sp. suggest that phosphorus availability has an impact on the degree of ploidy observed in this cyanobacterial species (Zerulla et al. 2016). Our data reveal higher polyploidy within phosphorous-replete cultures compared to 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 *Trichodesmium* 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


Table and figure legends

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

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

Table 3. Calculated potential contributions to total *in situ* chlorophyll by *Trichodesmium* spp. The calculated contribution that *Trichodesmium* makes to measured *in 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.
Table 1

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

Table 2

<table>
<thead>
<tr>
<th>Gene targets</th>
<th>Predicted ratio</th>
<th>Measured ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>nifH:16S rRNA</td>
<td>1:2</td>
<td>1:1.53 (±0.06)</td>
</tr>
<tr>
<td>nifH:psbAll</td>
<td>1:3</td>
<td>1:3.32 (±0.13)</td>
</tr>
</tbody>
</table>
### Table 3

*Trichodesmium* contribution to total chlorophyll (%) estimated from either cell counts or gene copies (assuming monoploidy)

<table>
<thead>
<tr>
<th>Range</th>
<th>Avg ± SD</th>
<th>Range</th>
<th>Avg ± SD</th>
<th>n</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-8.0</td>
<td>4.1 ± 1.7</td>
<td>NA</td>
<td>NA</td>
<td>16</td>
<td>This study, AMT 21</td>
</tr>
<tr>
<td>0-11.6</td>
<td>4.4 ± 7.9</td>
<td>NA</td>
<td>NA</td>
<td>33</td>
<td>Fernández et al. 2010**</td>
</tr>
<tr>
<td>0.2-13.7</td>
<td>7.2 ± 4.5</td>
<td>NA</td>
<td>NA</td>
<td>22</td>
<td>Letelier &amp; Karl 1996**</td>
</tr>
<tr>
<td>0-175.5</td>
<td>18.8 ± 30.3</td>
<td>NA</td>
<td>NA</td>
<td>336</td>
<td>Capone et al. 2004**</td>
</tr>
<tr>
<td>0-243.9</td>
<td>22.9 ± 34.0</td>
<td>NA</td>
<td>NA</td>
<td>335</td>
<td>Borstad 1978**</td>
</tr>
<tr>
<td>0-6</td>
<td>2.1 ± 2.5</td>
<td>0-98.4</td>
<td>10 ± 24.9</td>
<td>31</td>
<td>This study, AMT 17</td>
</tr>
<tr>
<td>0-17.8</td>
<td>4.1 ± 4.3</td>
<td>0-138.7</td>
<td>44.3 ± 63.9</td>
<td>15</td>
<td>This study, D361</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>1-440</td>
<td>136.0 ± 126.1</td>
<td>13</td>
<td>Goebel et al. 2010*</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>0-1163.8</td>
<td>99.6 ± 275.6</td>
<td>23</td>
<td>Kong et al. 2011**</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>0-2065.5</td>
<td>120.3 ± 319.1</td>
<td>125</td>
<td>Moisander et al. 2010**</td>
</tr>
</tbody>
</table>

*Data estimated from figures, **Data from Luo et al. (2012) dataset.*
Figure 1

A

![Graph showing the relationship between nifH copies/L and cells/L. The graph includes data points for D361, AMT17, and EM59101. The 1:1 line is also plotted.](image)

B

C

![Images B and C. B shows a close-up view of a cell culture, while C shows a different view with more cells.](image)
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).

<table>
<thead>
<tr>
<th>Homologues in</th>
<th>Gene</th>
<th>Primer</th>
<th>Sequence (5’-3’)</th>
<th>Amplicon size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T. erythraeum</td>
<td><em>nifH</em></td>
<td>Forward</td>
<td>TGGCCGTGGTATTATTACTGCT</td>
<td>111</td>
</tr>
<tr>
<td>IMS 101&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td>Reverse</td>
<td>ATC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Probe</td>
<td>GCAAATCCACCGCAACAAC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AAGGAGCTTATACAGATCTA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16S rRNA</td>
<td>Forward</td>
<td>CCACACTGGGACTGAGACAC</td>
<td>117</td>
</tr>
<tr>
<td></td>
<td>Tery_R0014, Tery_R0029</td>
<td>Reverse</td>
<td>AACCTAGAGCCTTCCTCCC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tery_0182, Tery_0183, Tery_4763</td>
<td>Forward</td>
<td>AATGCACCCATTCCACATGC</td>
<td>192</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reverse</td>
<td>CGACCGAAGTAGCCATGAG</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> annotated in the *Trichodesmium erythraeum* IMS 101 genome (Genbank accession NC_008312).