Supplement of

Version 2 of the global oceanic diazotroph database

Zhibo Shao¹,#, Yangchun Xu¹,#, Hua Wang¹, Weicheng Luo¹, Lice Wang¹, Yuhong Huang¹, Ya-Wei Luo¹

¹ State Key Laboratory of Marine Environmental Science and College of Ocean and Earth Sciences, Xiamen University, Xiamen, Fujian 361102, China

# These authors contributed equally.

Correspondence to: Ya-Wei Luo (ywlzuo@xmu.edu.cn)

Includes:

Fig. S1- Fig. S8
Table S1 & Table S2
References
Figure S1. Histogram of volumetric N\textsubscript{2} fixation rates measurements for (a) Trichodesmium N\textsubscript{2} fixation, (b) UCYN N\textsubscript{2} fixation, (c) heterocystous N\textsubscript{2} fixation, (d) whole seawater N\textsubscript{2} fixation and (e) total N\textsubscript{2} fixation (non-zero values). Data values are on logarithmic scale. Red line denotes the results of quality control, indicating the critical values above which data are likely outliers.
Figure S2. Histogram of depth-integrated N₂ fixation rates measurements for (a) *Trichodesmium* N₂ fixation, (b) UCYN N₂ fixation, (c) heterocystous N₂ fixation, (d) whole seawater N₂ fixation and (e) total N₂ fixation (non-zero values). Data values are on logarithmic scale.
Figure S3. Histogram of volumetric cell counts measurements for (a) *Trichodesmium*, (b) UCYN-A, (c) *Richelia*, and (d) *Calothrix* (non-zero values). Data values are on logarithmic scale. Red line denotes the results of quality control, indicating the critical values above which data are likely outliers.
Figure S4. Histogram of depth-integrated cell counts measurements for (a) *Trichodesmium* and (b) *Richelia* (non-zero values). Data values are on logarithmic scale.
Figure S5. Histogram of volumetric *nifH* qPCR copies measurements for (a) *Trichodesmium*, (b) UCYN-A, (c) UCYN-B, (d) UCYN-C (e) *Richelia*, and (f) *Calothrix* (non-zero values). Data values are on logarithmic scale. Red line denotes the results of quality control, indicating the critical values above which data are likely outliers.
Figure S6. Histogram of depth-integrated *nifH* qPCR copies measurements for (a) *Trichodesmium*, (b) UCYN-A, (c) UCYN-B, (d) UCYN-C (e) *Richelia*, and (f) *Calothrix* (non-zero values). Data values are on logarithmic scale. Red lines denote the results of quality control, indicating the critical values above which data are likely outliers.
Figure S7. Volumetric cell-count-based abundance of diazotrophic groups. The panels show *Trichodemium* (a–c), *Richelia* (d–f) and *Calothrix* (g–i) volumetric data in depth ranges of 25–100 m (a, d, g), 100-200 m (b, e, h) and below 200 m (c, f, i). For clear demonstration, data are binned to $3^\circ \times 3^\circ$ and geometric means in each bin are shown. Zero-value data are denoted as black crosses.
Figure S8. Volumetric nifH copies of different diazotrophic groups. The panels show Trichodemium (a–c), UCYN-A (d–f), UCYN-B (g–i), UCYN-C (j–l), Richelia (m–o) and Calothrix (p, q) volumetric data in depth ranges of 25–100 m (a, d, g, j, m, p), 100–200 m (b, e, h, k, n, q) and below 200 m (c, f, i, l, o). For clear demonstration, data are binned to 3° × 3° and geometric means in each bin are shown. Zero-value data are denoted as black crosses. Calothrix volumetric abundance has no data below 200 m.
Table S1. Summary of ratio of \textit{nifH} gene copies to cell numbers in published paper. The numbers in parentheses are ranges of observations.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Species</th>
<th>Mean cells L(^{-1})</th>
<th>Mean \textit{nifH} gene copies L(^{-1})</th>
<th>Mean \textit{nifH} copies cell(^{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>White et al. (2018)</td>
<td>\textit{Trichodesmium}</td>
<td>1,130 (8 – 4,130)</td>
<td>207,000 (593 – 1,460,000)</td>
<td>148</td>
</tr>
<tr>
<td></td>
<td>Het 1-3</td>
<td>190(^a) (78 – 390)</td>
<td>21,200 (148 – 124,000)</td>
<td>76</td>
</tr>
<tr>
<td>Lu et al. (2018)</td>
<td>\textit{Trichodesmium}</td>
<td>221,000 (19,000 – 423,000(^b))</td>
<td>2,020,000 (42,300 – 40,400,000(^b))</td>
<td>92</td>
</tr>
<tr>
<td>Sargent et al. (2016(^c))</td>
<td>\textit{Trichodesmium}</td>
<td>(56 – 18,000)(^d)</td>
<td>(0 – 270,000)(^d)</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0 – 19,000)(^d)</td>
<td>(0 – 670,000)(^d)</td>
<td>31</td>
</tr>
<tr>
<td>Krupke et al. (2013)</td>
<td>UCYN-A</td>
<td>134,000 (1,760 – 819,000)</td>
<td>746,000 (37,200 – 3,690,000)</td>
<td>14</td>
</tr>
</tbody>
</table>

\(^a\) Heterocystous cell numbers were reported, here total cell numbers were used (heterocystous cell numbers multiply a factor of 6 assuming one filament contains one heterocystous cell and five vegetative cells).
\(^b\) data collected during \textit{Trichodesmium} bloom.
\(^c\) two cruises in different seasons.
\(^d\) no average data reported.
**Table S2.** Biomass conversion factors for heterocystous diazotroph-diatom associations. The numbers in parentheses are ranged used in the estimation.

<table>
<thead>
<tr>
<th>Species</th>
<th>Diatom Biovolume</th>
<th>Biomass&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Vegetative cells per heterocyst</th>
<th>Heterocysts per DDA</th>
<th>Number of diazotrophic cells per DDA&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Diazotrophic cell biomass conversion&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Total DDA biomass&lt;sup&gt;c&lt;/sup&gt;</th>
<th>DDA Biomass Conversion factor&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Richelia-Hemiaulus</td>
<td>10800 (7050–20900)</td>
<td>418 (287–748)</td>
<td>2</td>
<td>12 (8–22)</td>
<td>144 (16–1760)</td>
<td>562 (303–2508)</td>
<td>281 (152–1254)</td>
<td></td>
</tr>
<tr>
<td>Calothrix-Chaetoceros</td>
<td>1480 (300–4630)</td>
<td>73 (18–288)</td>
<td>2</td>
<td>12 (8–22)</td>
<td>144 (40–440)</td>
<td>217 (58–728)</td>
<td>110 (29–364)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Based on an empirical relationship (Menden-Deuer and Lessard, 2000): biomass = 0.117 × biovolume<sup>0.881</sup>.<br><sup>b</sup> (1 + number of vegetative cells) × number of heterocyst.<br><sup>c</sup> Diatom biomass + diazotrophic biomass.
Reference


