Supplement of Earth Syst. Sci. Data, 15, 3673–3709, 2023
https://doi.org/10.5194/essd-15-3673-2023-supplement
© Author(s) 2023. CC BY 4.0 License.

Supplement of

Global oceanic diazotroph database version 2 and elevated estimate of global oceanic N$_2$ fixation

Zhibo Shao et al.

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

The copyright of individual parts of the supplement might differ from the article licence.
Supporting information

Includes:

Fig. S1- Fig. S7
Table S1 - Table S4
References

Figure S1. Histogram of volumetric N₂ fixation rates measurements for (a) heterocystous N₂ fixation, (b) <10 μm diazotrophic N₂ fixation, (c) >10 μm diazotrophic N₂ fixation, (d) whole seawater N₂ fixation, (e) total N₂ fixation and (f) depth-integrated N₂ 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 volumetric cell counts measurements for (a) *Trichodesmium*, (b) Het-1+Het-2, and (c) Het-3 (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 S3. Histogram of depth-integrated cell counts measurements for (a) *Trichodesmium* and (b) Het-1 + Het-2 (non-zero values). Data values are on logarithmic scale.
Figure S4. Histogram of volumetric nifH qPCR copies measurements for (a) *Trichodesmium*, (b) UCYN-A, (c) UCYN-B, (d) UCYN-C (e) Het-1 + Het-2, and (f) Het-3 (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 S5. Histogram of depth-integrated nifH qPCR copies measurements for (a) *Trichodesmium*, (b) UCYN-A, (c) UCYN-B, (d) UCYN-C (e) Het-1 + Het-2, and (f) Het-3 (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 S6. Volumetric cell abundance of diazotrophic groups. The panels show *Trichodemium* (a–c), Het-1 + Het-2 (d–f) and Het-3 (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 open black circles.
Figure S7. Volumetric $nifH$ gene copy abundances of different diazotrophic groups. The panels show *Trichodemium* (a–c), UCYN-A (d–f), UCYN-B (g–i), UCYN-C (j–l), Het-1 + Het-2 (m–o) and Het-3 (p, q) volumetric data in depth ranges of 25–100 m (a, d, g, j, m, p), 100–200 (b, e, h, k, n, q) and below 200 m (c, f, i, l, o). 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 open black circles. *Calothrix* volumetric abundance has no data below 200 m.
**Table S1.** Summary of data points of cell-specific N₂ fixation rates in the version 2 of the database, including measurements for *Trichodesmium*, UCYN-A, UCYN-A1, UCYN-A2, UCYN-B, *Richelia*, *Calothrix*, cyanobacteria, and non-cyanobacteria diazotrophs (NCDs).

<table>
<thead>
<tr>
<th>Reference</th>
<th>Region</th>
<th><em>Trichodesmium</em></th>
<th>UCYN-A</th>
<th>UCYN-A1</th>
<th>UCYN-A2</th>
<th>UCYN-B</th>
<th>Richelia</th>
<th>Calothrix</th>
<th>Unclassified Cyano</th>
<th>NCDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benavides et al.</td>
<td>Tropical SW Pacific</td>
<td>101.65 ± 25.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2022)</td>
<td></td>
<td>(n = 6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benavides et al.</td>
<td>SW Pacific</td>
<td>277.32 ± 189.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2017)</td>
<td></td>
<td>(n = 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bonnet et al. (2018)</td>
<td>Tropical SW Pacific</td>
<td>64.00 ± 30.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(n = 3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foster et al. (2011)</td>
<td>N Pacific</td>
<td>91.80 ± 52.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(n = 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foster et al. (2013)</td>
<td>N Pacific</td>
<td>11.64 ± 2.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(n = 6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foster et al. (2022)</td>
<td>Tropical NW Atlantic</td>
<td>8.07 ± 1.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(n = 5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gradoville et al.</td>
<td>N Pacific</td>
<td>3.34 ± 0.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2021)</td>
<td></td>
<td>(n = 17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harding et al. (2018)</td>
<td>Arctic Ocean</td>
<td>7.10 ± 6.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(n = 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harding et al. (2022)</td>
<td>Subtropical N Pacific</td>
<td>0.33 ± 0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(n = 40)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Krupke et al. (2013)</td>
<td>Subtropical N Atlantic</td>
<td>0.04 ± 0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(n = 4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.78 ± 4.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Rates in unit of fmol N cell⁻¹ d⁻¹
<table>
<thead>
<tr>
<th>Reference</th>
<th>Region</th>
<th>Trichodesmium</th>
<th>UCYN-A</th>
<th>UCYN-A1</th>
<th>UCYN-A2</th>
<th>UCYN-B</th>
<th>Richelia</th>
<th>Calothrix</th>
<th>Unclassified Cyanobacteria</th>
<th>NCDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Krupke et al. (2015)</td>
<td>Subtropical NE Atlantic</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(n = 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martínez-Pérez et al. (2016)</td>
<td>Tropical N Atlantic</td>
<td>51.03 ± 6.41</td>
<td>3.39 ± 0.44</td>
<td>172.26 ± 12.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(n = 101)</td>
<td>(n = 57)</td>
<td>(n = 10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McCarthy and Carpenter (1979)</td>
<td>N Atlantic</td>
<td>31.72 ± 4.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(n = 24)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mills et al. (2020)</td>
<td>California Current System</td>
<td>0.51 ± 0.07</td>
<td>5.96 ± 1.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(n = 15)</td>
<td>(n = 9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turk-Kubo et al. (2021)</td>
<td>Southern California Current System</td>
<td>1.12 ± 0.41</td>
<td>17.45 ± 2.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(n = 26)</td>
<td>(n = 17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td></td>
<td><strong>53 ± 6</strong></td>
<td><strong>0.05 ± 0.01</strong></td>
<td><strong>2.7 ± 0.3</strong></td>
<td><strong>55 ± 12</strong></td>
<td><strong>26 ± 11</strong></td>
<td><strong>216 ± 44</strong></td>
<td><strong>860 ± 856</strong></td>
<td><strong>0.33 ± 0.05</strong></td>
<td><strong>0.69 ± 0.27</strong></td>
</tr>
</tbody>
</table>

*Rates in a different unit (d⁻¹)*

<table>
<thead>
<tr>
<th>Reference</th>
<th>Region</th>
<th>Trichodesmium</th>
<th>UCYN-A</th>
<th>UCYN-A1</th>
<th>UCYN-A2</th>
<th>UCYN-B</th>
<th>Richelia</th>
<th>Calothrix</th>
<th>Unclassified Cyanobacteria</th>
<th>NCDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filella et al. (2022)</td>
<td>S Pacific Gyre</td>
<td>0.02 ± 0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.00 ± 0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(n = 13)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(n = 12)</td>
</tr>
</tbody>
</table>
Table S2. Summary of comparison measurements of abundance of cells and *nifH* gene copies, including reported mean abundance and ranges (numbers in parentheses). The ratios of *nifH* gene copies to cell abundance are shown as mean ± s.d.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Groups</th>
<th>Cell abundance (cells L(^{-1}))</th>
<th><em>NifH</em> abundance (gene copies L(^{-1}))</th>
<th><em>NifH</em>:cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gradoville et al. (2022)</td>
<td><em>Crocosphaera</em></td>
<td>89,900 (220 – 240,000)</td>
<td>165,000 (290 – 378,000)</td>
<td>1.8(^a)</td>
</tr>
<tr>
<td>Gradoville et al. (2022)</td>
<td><em>Trichodesmium</em></td>
<td>2,340 (235 – 5,900)</td>
<td>15,300 (775 – 145,000)</td>
<td>2.7(^a)</td>
</tr>
<tr>
<td>Gradoville et al. (2022)</td>
<td><em>Richelia</em></td>
<td>950 (160 – 2,330)</td>
<td>2,630 (260 – 5,230)</td>
<td>2.6(^a)</td>
</tr>
<tr>
<td>Gradoville et al. (2022)</td>
<td><em>Calothrix</em></td>
<td>1,020 (156 – 2,000)</td>
<td>1,400 (136 – 2,390)</td>
<td>1.5(^a)</td>
</tr>
<tr>
<td>White et al. (2018)</td>
<td><em>Trichodesmium</em></td>
<td>1,130 (8 – 4,130)</td>
<td>207,000 (593 – 1,460,000)</td>
<td>148 ± 248</td>
</tr>
<tr>
<td>White et al. (2018)</td>
<td>Het 1-3</td>
<td>190(^b) (78 – 390)</td>
<td>21,200 (148 – 124,000)</td>
<td>76 ± 93</td>
</tr>
<tr>
<td>Lu et al. (2018)</td>
<td><em>Trichodesmium</em></td>
<td>221,000 (19,000 – 423,000)</td>
<td>2,020,000 (42,300 – 40,400,000)</td>
<td>49 ± 66</td>
</tr>
<tr>
<td>Wilson et al. (2017)</td>
<td><em>Crocosphaera</em></td>
<td>500</td>
<td>1800</td>
<td>3.6</td>
</tr>
<tr>
<td>Sargent et al. (2016)</td>
<td><em>Trichodesmium</em></td>
<td>(56 – 18,000)(^d)</td>
<td>(0 – 270,000)(^d)</td>
<td>12 ± 13</td>
</tr>
<tr>
<td>Sargent et al. (2016)</td>
<td><em>Trichodesmium</em></td>
<td>(0 – 19,000)(^d)</td>
<td>(0 – 670,000)(^d)</td>
<td>31 ± 30</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 ± 10</td>
</tr>
</tbody>
</table>

\(^a\) Slope in the regression test.

\(^b\) Reported heterocystous cell numbers are converted to total diazotrophic cells by assuming each heterocystous cell is associated with five vegetative cells.

\(^c\) Data collected during *Trichodesmium* blooms.

\(^d\) No average data reported.
Table S3. Summary of a few considerations for application and interpretation of qPCR and microscopy counting for enumeration and activity (RNA) of diazotrophs.

<table>
<thead>
<tr>
<th>Consideration</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell identity</td>
<td><em>Microscopy</em>: Cross-comparison of cell counts can be difficult as training and experience varies.</td>
</tr>
<tr>
<td>Patchy distribution, low abundance</td>
<td><em>Both methods</em>: collection of samples (volumes, depths) are dependent on logistics; collection strategies can vary: size fractionation, gravity filtration (microscopy), etc. <em>Microscopy</em>: potential to underestimate/overestimate if the minimum number of cells is not enumerated. <em>qPCR</em>: potential to underestimate if targets are below detection of assay (1-10 copies).</td>
</tr>
<tr>
<td>Dead or moribund cells</td>
<td><em>Both methods</em> do not distinguish vitality, thus potential to overestimate.</td>
</tr>
<tr>
<td>Primer design</td>
<td><em>qPCR</em>: potential to overestimate if primers cross-react with non-targets; potential to underestimate if primers are too specific to a limited/unknown micro-diversity.</td>
</tr>
<tr>
<td>Polyploidy</td>
<td>Some bacteria, including data on <em>Trichodesmium</em>, <em>Richelia</em>, generate multiple genome copies during their life cycle (Sargent et al., 2016; White et al., 2018; Karlush et al., 2021).</td>
</tr>
<tr>
<td>Gene copy number</td>
<td>Filamentous cyanobacteria (includes heterocystous cyanobacteria) possess a genome copy in each cell; it is not known for all diazotrophs the number of <em>nifH</em> copies/cell, often assumed to be one.</td>
</tr>
<tr>
<td>DNA/RNA Extraction efficiency</td>
<td>Not all targets extract uniformly; RNA is prone to degrade</td>
</tr>
</tbody>
</table>
### Table S4. Biomass conversion factors for heterocystous diazotroph-diatom associations. The numbers in parentheses are ranged used in the estimation.

<table>
<thead>
<tr>
<th>Richelia Groups</th>
<th>Richelia-Hemiaulus</th>
<th>Richelia-Rhizosolenia</th>
<th>Richelia-Chaetoceros</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of observations</strong></td>
<td>251</td>
<td>92</td>
<td>48</td>
</tr>
<tr>
<td><strong>Heterocyst cell</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biovolume (μm³)</td>
<td>97 (34−678)</td>
<td>248 (29−1124)</td>
<td>62 (17−258)</td>
</tr>
<tr>
<td>Biomass (pg C cell⁻¹)</td>
<td>22 (9−120)</td>
<td>51 (8−186)</td>
<td>15 (5−52)</td>
</tr>
<tr>
<td><strong>Vegetative cell</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biovolume (μm³)</td>
<td>14 (2−87)</td>
<td>31 (4−151)</td>
<td>14 (6−31)</td>
</tr>
<tr>
<td>Biomass (pg C cell⁻¹)</td>
<td>4 (1−20)</td>
<td>8 (2−33)</td>
<td>4 (2−8)</td>
</tr>
<tr>
<td><strong>Diazotrophic filaments</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterocysts per DDA</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Vegetative cells per heterocyst</td>
<td>3 (1−8)</td>
<td>5 (3−10)</td>
<td>5 (1−19)</td>
</tr>
<tr>
<td>Number of vegetative cells per DDA</td>
<td>3 (1−8)</td>
<td>6 (2−16)</td>
<td>5 (3−10)</td>
</tr>
<tr>
<td>Diazotrophic biomass (pg C DDA⁻¹)</td>
<td>34 (10−280)</td>
<td>68 (20−560)</td>
<td>91 (14−516)</td>
</tr>
<tr>
<td><strong>Diatom</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biovolume (μm³)</td>
<td>10800 (7050−20900)</td>
<td>57100 (439−188000)</td>
<td>1480 (300−4630)</td>
</tr>
<tr>
<td>Biomass (pg C cell⁻¹)</td>
<td>418 (287−748)</td>
<td>1814 (25−5184)</td>
<td>73 (18−288)</td>
</tr>
<tr>
<td><strong>DDA</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biomass (pg C heterocyst⁻¹)</td>
<td>452 (297−1028)</td>
<td>243 (154−654)</td>
<td>1905 (39−5700)</td>
</tr>
</tbody>
</table>

a number of vegetative cells × vegetative cell biomass + number of heterocyst × heterocyst cell biomass.
b Based on an empirical relationship (Menden-Deuer and Lessard, 2000): biomass = 0.117 × biovolume⁰.⁸⁸¹.
Biovolume of diatom comes from a database (Harrison et al., 2015)
c (Diatom biomass + diazotrophic biomass)/number of heterocyst.

a number of vegetative cells × vegetative cell biomass + number of heterocyst × heterocyst cell biomass.
b Based on an empirical relationship (Menden-Deuer and Lessard, 2000): biomass = 0.117 × biovolume⁰.⁸⁸¹.
Biovolume of diatom comes from a database (Harrison et al., 2015)
c (Diatom biomass + diazotrophic biomass)/number of heterocyst.
References


White, A. E., Watkins-Brandt, K. S., and Church, M. J.: Temporal variability of 
Trichodesmium spp. and diatom-diazotroph assemblages in the North Pacific subtropical 

Coordinated regulation of growth, activity and transcription in natural populations of the 
unicellular nitrogen-fixing cyanobacterium Crocosphaera, Nat. Microbiol, 2, 17118, 