



Supplement of

Global oceanic diazotroph database version 2 and elevated estimate of global oceanic N₂ 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

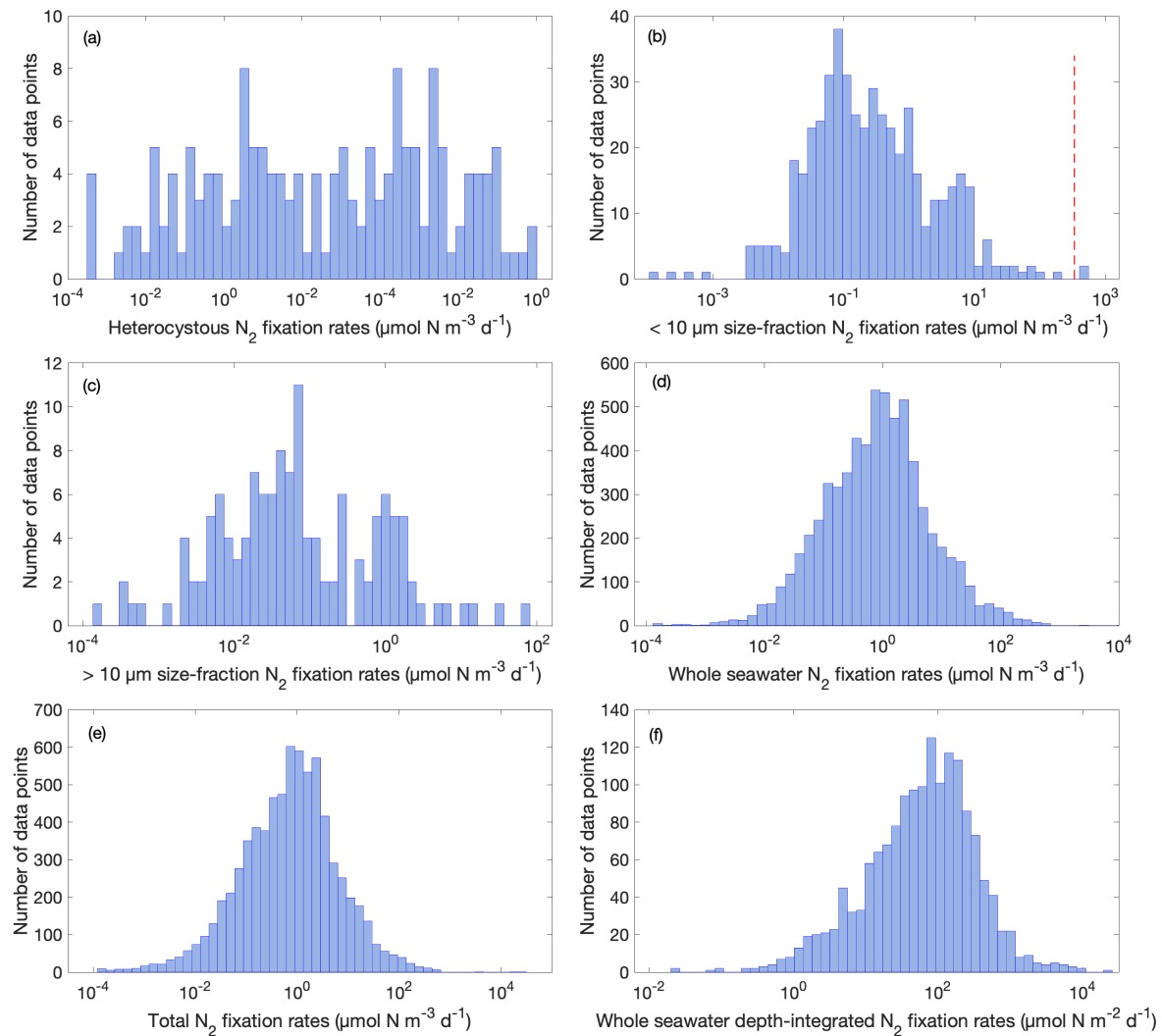


Figure S1. Histogram of volumetric N_2 fixation rates measurements for (a) heterocystous N_2 fixation, (b) $<10 \mu\text{m}$ diazotrophic N_2 fixation, (c) $>10 \mu\text{m}$ diazotrophic N_2 fixation, (d) whole seawater N_2 fixation, (e) total N_2 fixation and (f) depth-integrated N_2 fixation (non-zero 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.

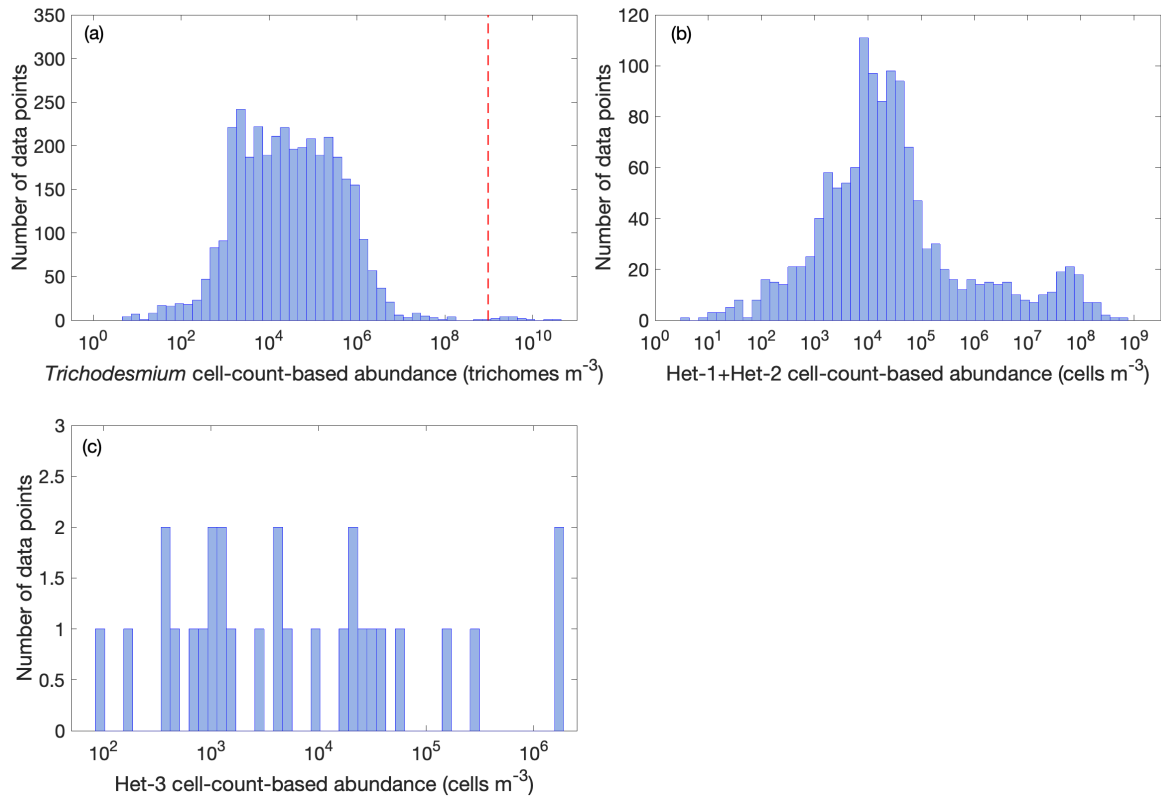


Figure S2. Histogram of volumetric cell counts measurements for (a) *Trichodesmium*, (b) Het-1+Het-2, and (c) Het-3 (non-zero values). Data values are on logarithmic scale. Red line denotes the results of quality control, indicating the critical values above which data are likely outliers.

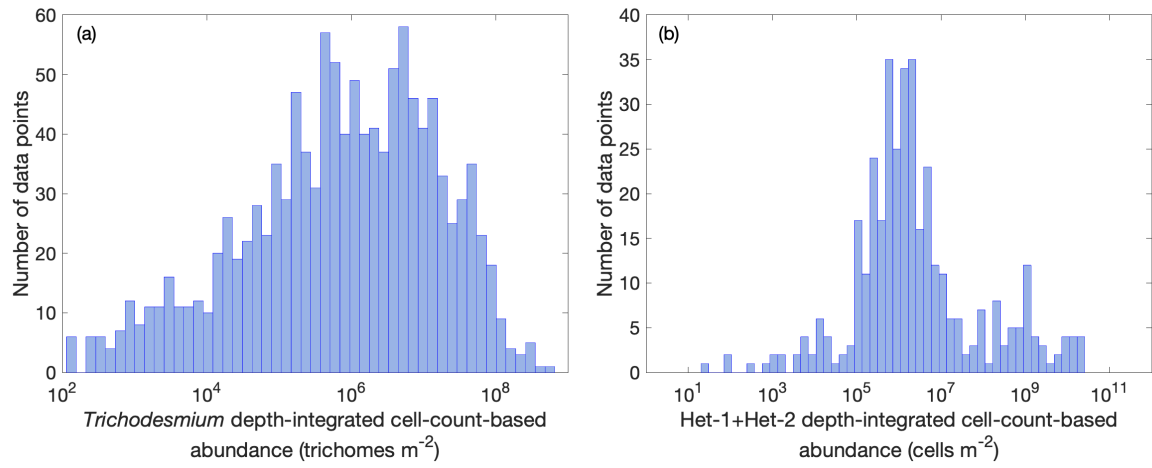


Figure S3. Histogram of depth-integrated cell counts measurements for (a) *Trichodesmium* and (b) Het-1 + Het-2 (non-zero values). Data values are on logarithmic scale.

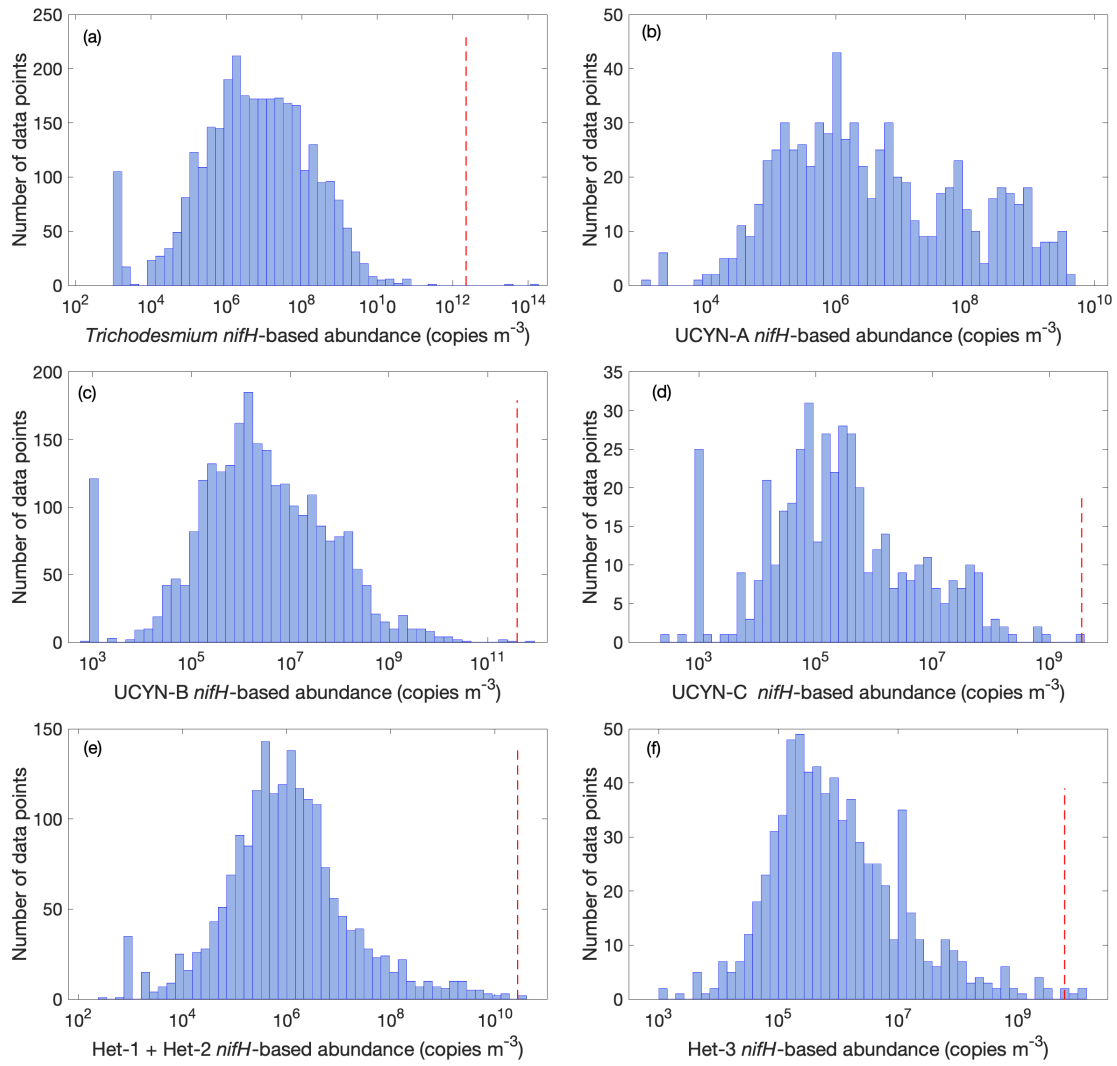


Figure S4. Histogram of volumetric *nifH* qPCR copies measurements for (a) *Trichodesmium*, (b) UCYN-A, (c) UCYN-B, (d) UCYN-C (e) Het-1 + Het-2, and (f) Het-3 (non-zero values). Data values are on logarithmic scale. Red line denotes the results of quality control, indicating the critical values above which data are likely outliers.

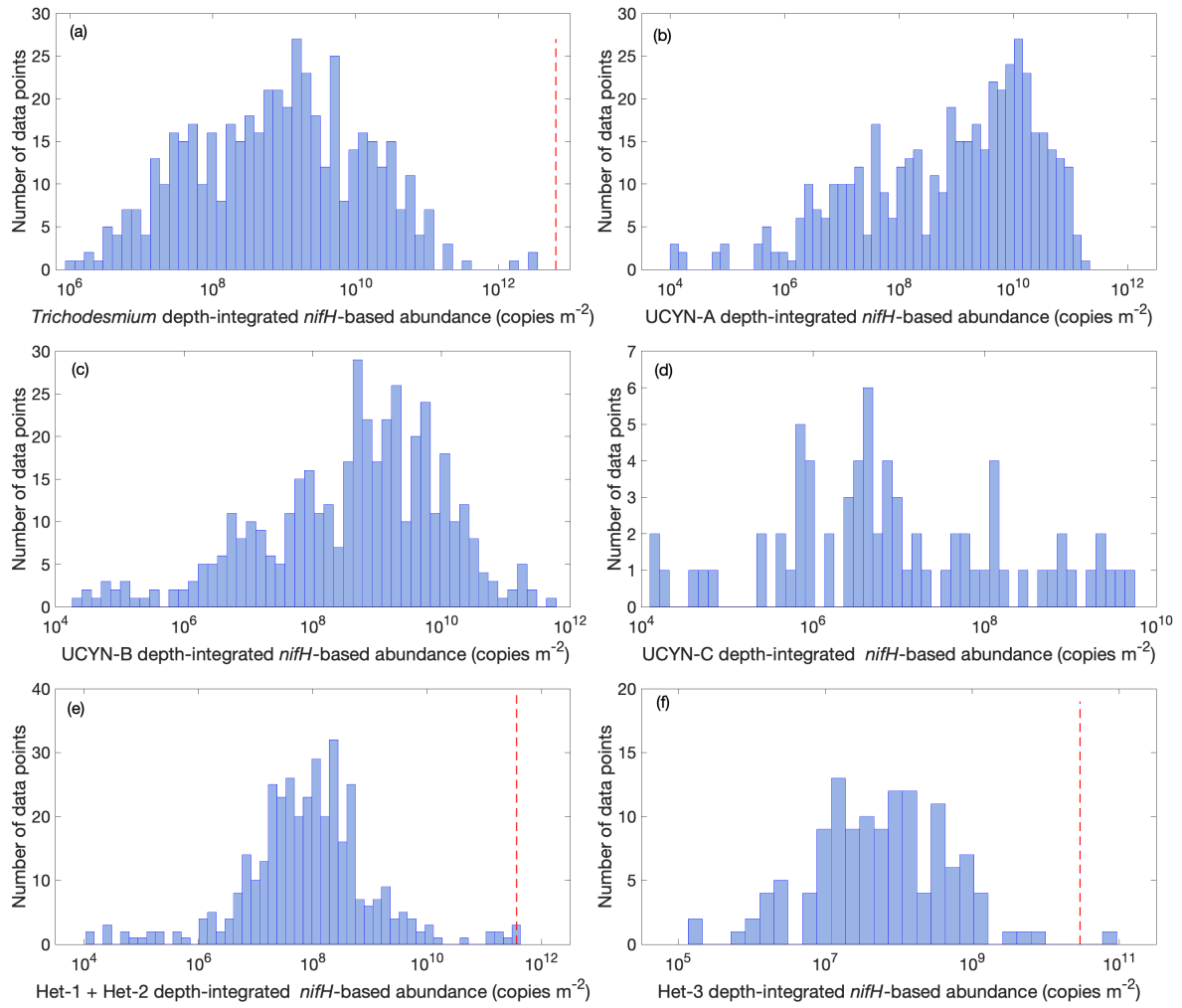


Figure S5. Histogram of depth-integrated *nifH* qPCR copies measurements for (a) *Trichodesmium*, (b) UCYN-A, (c) UCYN-B, (d) UCYN-C (e) Het-1 + Het-2, and (f) Het-3 (non-zero values). Data values are on logarithmic scale. Red lines denote the results of quality control, indicating the critical values above which data are likely outliers.

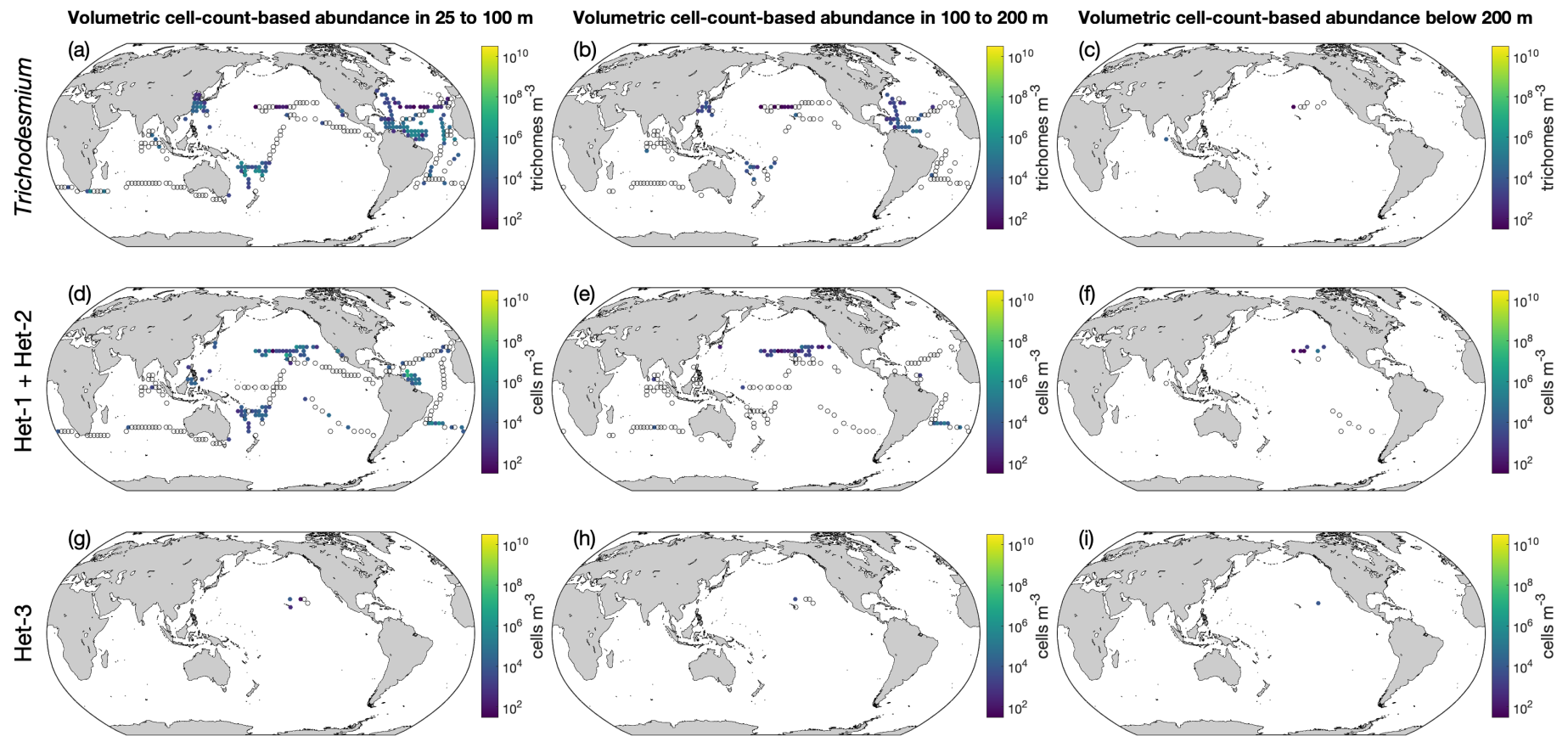


Figure S6. Volumetric cell abundance of diazotrophic groups. The panels show *Trichodesmium* (a–c), Het-1 + Het-2 (d–f) and Het-3(g–i) volumetric data in depth ranges of 25–100 m (a, d, g), 100–200 m (b, e, h) and below 200 m (c, f, i). For clear demonstration, data are binned to $3^\circ \times 3^\circ$ and geometric means in each bin are shown. Zero-value data are denoted as open black circles.

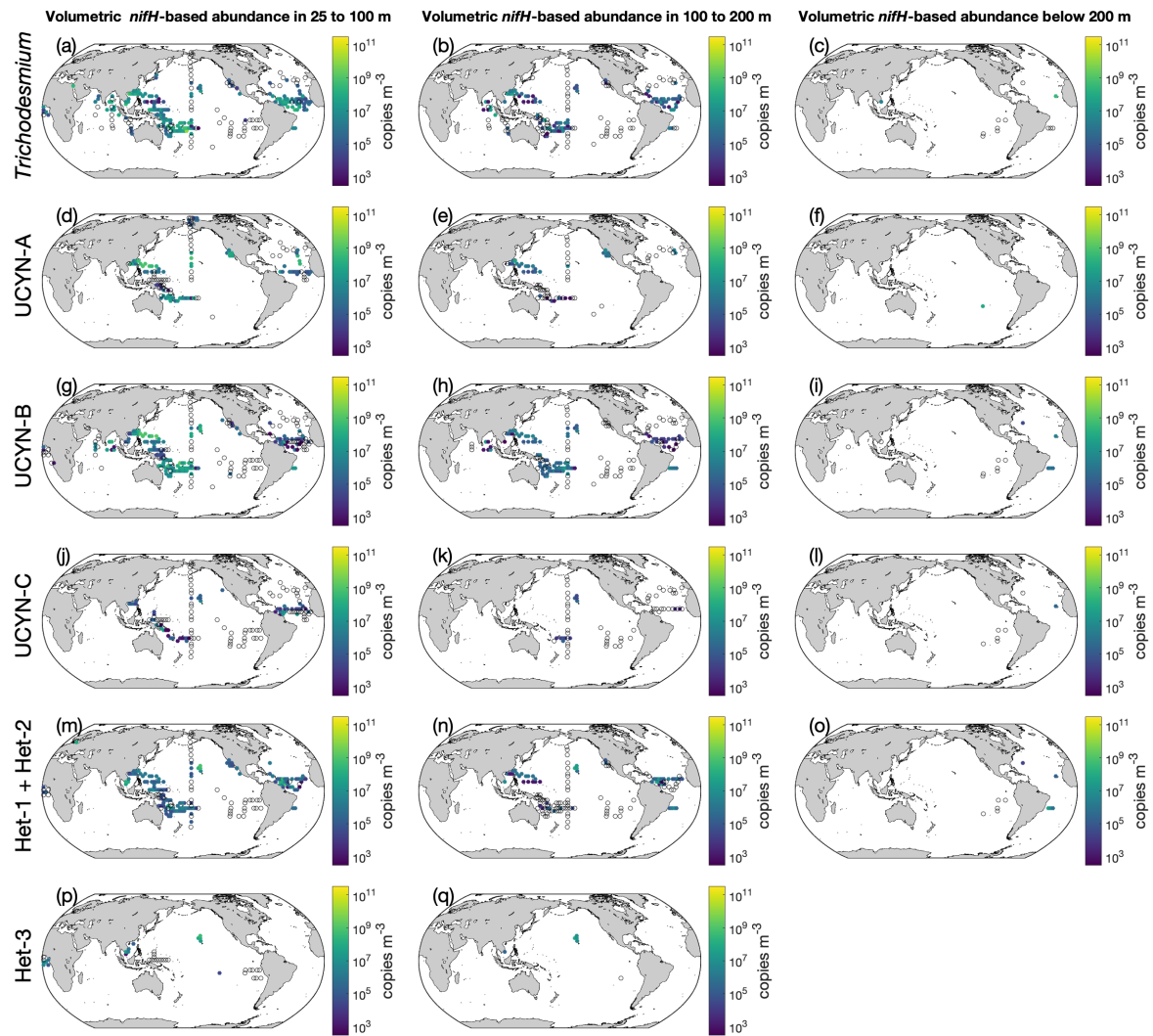


Figure S7. Volumetric *nifH* gene copy abundances of different diazotrophic groups. The panels show *Trichodesmium* (a–c), UCYN-A (d–f), UCYN-B (g–i), UCYN-C (j–l), Het-1 + Het-2 (m–o) and Het-3 (p, q) volumetric data in depth ranges of 25–100 m (a, d, g, j, m, p), 100–200 m (b, e, h, k, n, q) and below 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).

Reference	Region	<i>Trichodesmium</i>	UCYN-A	UCYN-A1	UCYN-A2	UCYN-B	<i>Richelia</i>	<i>Calothrix</i>	Unclassified Cyanobacteria	NCDs
<i>Rates in unit of fmol N cell⁻¹ d⁻¹</i>										
Benavides et al. (2022)	Tropical SW Pacific	101.65 ± 25.87 (n = 6)								
Benavides et al. (2017)	SW Pacific	277.32 ± 189.80 (n = 2)								
Bonnet et al. (2018)	Tropical SW Pacific	64.00 ± 30.03 (n = 3)				18.05 ± 11.95 (n = 2)				
Foster et al. (2011)	N Pacific					91.80 ± 52.92 (n = 2)	463.71 ± 105.08 (n = 18)	860.04 ± 855.96 (n = 2)		
Foster et al. (2013)	N Pacific					11.64 ± 2.35 (n = 6)				
Foster et al. (2022)	Tropical NW Atlantic						101.35 ± 29.09 (n = 39)			
Gradoville et al. (2020)	N Pacific			8.07 ± 1.11 (n = 5)						
Gradoville et al. (2021)	N Pacific			3.34 ± 0.44 (n=17)						
Harding et al. (2018)	Arctic Ocean				7.10 ± 6.00 (n = 2)					
Harding et al. (2022)	Subtropical N Pacific								0.33 ± 0.05 (n= 40)	0.69 ± 0.27 (n = 34)
Krupke et al. (2013)	Subtropical N Atlantic		0.04 ± 0.00 (n = 4)			8.78 ± 4.63 (n = 2)				

Reference	Region	<i>Trichodesmium</i>	UCYN-A	UCYN-A1	UCYN-A2	UCYN-B	<i>Richelia</i>	<i>Calothrix</i>	Unclassified Cyanobacteria	NCDs
Krupke et al. (2015)	Subtropical NE Atlantic		0.07 (n = 1)							
Martínez-Pérez et al. (2016)	Tropical N Atlantic	51.03 ± 6.41 (n = 101)		3.39 ± 0.44 (n = 57)	172.26 ± 12.88 (n = 10)					
McCarthy and Carpenter (1979)	N Atlantic	31.72 ± 4.85 (n = 24)								
Mills et al. (2020)	California Current System			0.51 ± 0.07 (n = 15)	5.96 ± 1.77 (n = 9)					
Turk-Kubo et al. (2021)	Southern California Current System			1.12 ± 0.41 (n = 26)	17.45 ± 2.73 (n = 17)					
Mean		53 ± 6	0.05 ± 0.01	2.7 ± 0.3	55 ± 12	26 ± 11	216 ± 44	860 ± 856	0.33 ± 0.05	0.69 ± 0.27
<i>Rates in a different unit (d⁻¹)</i>										
Filella et al. (2022)	S Pacific Gyre	0.02 ± 0.01 (n = 13)				0.00 ± 0.00 (n = 12)				

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 \pm s.d.

Reference	Groups	Cell abundance (cells L ⁻¹)	<i>NifH</i> abundance (gene copies L ⁻¹)	<i>NifH</i> :cell
Gradoville et al. (2022)	<i>Crocospaera</i>	89,900 (220 – 240,000)	165,000 (290 – 378,000)	1.8 ^a
Gradoville et al. (2022)	<i>Trichodesmium</i>	2,340 (235 – 5,900)	15,300 (775– 145,000)	2.7 ^a
Gradoville et al. (2022)	<i>Richelia</i>	950 (160 – 2,330)	2,630 (260 – 5,230)	2.6 ^a
Gradoville et al. (2022)	<i>Calothrix</i>	1,020 (156– 2,000)	1,400 (136– 2,390)	1.5 ^a
White et al. (2018)	<i>Trichodesmium</i>	1,130 (8 – 4,130)	207,000 (593 – 1,460,000)	148 \pm 248
White et al. (2018)	Het 1-3	190 ^b (78 – 390)	21,200 (148 – 124,000)	76 \pm 93
Lu et al. (2018)	<i>Trichodesmium</i>	221,000 (19,000 – 423,000 ^c)	2,020,000 (42,300 – 40,400,000 ^c)	49 \pm 66
Wilson et al. (2017)	<i>Crocospaera</i>	500	1800	3.6
Sargent et al. (2016)	<i>Trichodesmium</i>	(56 – 18,000) ^d	(0 – 270,000) ^d	12 \pm 13
Sargent et al. (2016)	<i>Trichodesmium</i>	(0 – 19,000) ^d	(0 – 670,000) ^d	31 \pm 30
Krupke et al. (2013)	UCYN-A	134,000 (1,760 – 819,000)	746,000 (37,200 – 3,690,000)	14 \pm 10

^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.

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

Table S4. Biomass conversion factors for heterocystous diazotroph-diatom associations. The numbers in parentheses are ranged used in the estimation.

<i>Richelia</i> Groups		<i>Richelia-Hemiaulus</i>		<i>Richelia-Rhizosolenia</i>		<i>Richelia-Chaetoceros</i>	
Number of observations		251		92		48	
Heterocyst cell	Biovolume (μm^3)	97 (34–678)		248 (29–1124)		62 (17–258)	
	Biomass (pg C cell^{-1})	22 (9–120)		51 (8–186)		15 (5–52)	
Vegetative cell	Biovolume (μm^3)	14 (2–87)		31 (4–151)		14 (6–31)	
	Biomass (pg C cell^{-1})	4 (1–20)		8 (2–33)		4 (2–8)	
Diazotrophic filaments	Heterocysts per DDA	1	2	1	5	3	10
	Vegetative cells per heterocyst	3 (1–8)		5 (3–10)		5 (1–19)	
	Number of vegetative cells per DDA	3 (1–8)	6 (2–16)	5 (3–10)	25 (15–50)	15 (3–57)	50 (10–190)
	Diazotrophic biomass (pg C DDA^{-1}) ^a	34 (10–280)	68 (20–560)	91 (14–516)	455 (70–2580)	105 (21–612)	350 (70–2040)
Diatom	Biovolume (μm^3)	10800 (7050–20900)		57100 (439–188000)		1480 (300–4630)	
	Biomass (pg C cell^{-1}) ^b	418 (287–748)		1814 (25–5184)		73 (18–288)	
DDA	Biomass ($\text{pg C heterocyst}^{-1}$) ^c	452 (297–1028)	243 (154–654)	1905 (39–5700)	454 (19–1553)	59 (13–300)	42 (9–233)

^a number of vegetative cells \times vegetative cell biomass + number of heterocyst \times heterocyst cell biomass.

^b Based on an empirical relationship (Menden-Deuer and Lessard, 2000) : biomass = $0.117 \times \text{biovolume}^{0.881}$.
Biovolume of diatom comes from a database (Harrison et al., 2015)

^c (Diatom biomass + diazotrophic biomass)/number of heterocyst.

^a number of vegetative cells \times vegetative cell biomass + number of heterocyst \times heterocyst cell biomass.

^b Based on an empirical relationship (Menden-Deuer and Lessard, 2000) : biomass = $0.117 \times \text{biovolume}^{0.881}$.
Biovolume of diatom comes from a database (Harrison et al., 2015)

^c (Diatom biomass + diazotrophic biomass)/number of heterocyst.

References

- Foster, R. A., Kuypers, M. M. M., Vagner, T., Paerl, R. W., Musat, N., and Zehr, J. P.: Nitrogen fixation and transfer in open ocean diatom–cyanobacterial symbioses, *The ISME Journal*, 5, 1484-1493, doi:<https://doi.org/10.1038/ismej.2011.26>, 2011.
- Gradoville, M., Cabello, A., Wilson, S., Turk-Kubo, K., Karl, D., and Zehr, J.: Light and depth dependency of nitrogen fixation by the non-photosynthetic, symbiotic cyanobacterium UCYN-A, *Environmental microbiology*, 23, doi:<https://doi.org/10.1111/1462-2920.15645>, 2021.
- Gradoville, M. R., Farnelid, H., White, A. E., Turk-Kubo, K. A., Stewart, B., Ribalet, F., Ferron, S., Pinedo-Gonzalez, P., Armbrust, E. V., Karl, D. M., John, S., and Zehr, J. P.: Latitudinal constraints on the abundance and activity of the cyanobacterium UCYN-A and other marine diazotrophs in the North Pacific, *Limnol Oceanogr*, doi:<https://doi.org/10.1002/lno.11423>, 2020.
- Gradoville, M. R., Dugenne, M., Hynes, A. M., Zehr, J. P., and White, A. E.: Empirical relationship between *nifH* gene abundance and diazotroph cell concentration in the North Pacific Subtropical Gyre, *Journal of Phycology*, doi:<https://doi.org/10.1111/jpy.13289>, 2022.
- Harrison, P., Zingone, A., Mickelson, M., Lehtinen, S., Nagappa, R., Kraberg, A., Sun, J., McQuatters-Gollop, A., and Jakobsen, H.: Cell volumes of marine phytoplankton from globally distributed coastal data sets, *Estuarine Coastal Shelf Sci.*, 162, doi:<https://doi.org/10.1016/j.ecss.2015.05.026>, 2015.
- Karlusich, J. J. P., Pelletier, E., Lombard, F., Carsique, M., Dvorak, E., Colin, S., Picheral, M., Cornejo-Castillo, F. M., Acinas, S. G., Pepperkok, R., Karsenti, E., De Vargas, C., Wincker, P., Bowler, C., and Foster, R. A.: Global distribution patterns of marine nitrogen-fixers by imaging and molecular methods, *Nat. Commun.*, 12, doi:<https://doi.org/10.1038/s41467-021-24299-y>, 2021.
- Krupke, A., Musat, N., LaRoche, J., Mohr, W., Fuchs, B. M., Amann, R. I., Kuypers, M. M. M., and Foster, R. A.: In situ identification and N₂ and C fixation rates of uncultivated cyanobacteria populations, *Syst. Appl. Microbiol.*, 36, 259-271, doi:<https://doi.org/10.1016/j.syapm.2013.02.002>, 2013.
- Lu, Y., Wen, Z., Shi, D., Chen, M., Zhang, Y., Bonnet, S., Li, Y., Tian, J., and Kao, S. J.: Effect of light on N₂ fixation and net nitrogen release of *Trichodesmium* in a field study, *Biogeosciences*, 15, 1-12, doi:<https://doi.org/10.5194/bg-15-1-2018>, 2018.
- Menden-Deuer, S., and Lessard, E. J.: Carbon to volume relationships for dinoflagellates, diatoms, and other protist plankton, *Limnol Oceanogr*, 45, 569-579, doi:<https://doi.org/10.4319/lo.2000.45.3.0569>, 2000.
- Sargent, E. C., Hitchcock, A., Johansson, S. A., Langlois, R., Moore, C. M., LaRoche, J., Poulton, A. J., and Bibby, T. S.: Evidence for polyploidy in the globally important diazotroph *Trichodesmium*, *FEMS Microbiol. Lett.*, 363, fnw244, doi:<https://doi.org/10.1093/femsle/fnw244>, 2016.

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 gyre, *Front. Mar. Sci.*, 5, 27, doi:<https://doi.org/10.3389/fmars.2018.00027>, 2018.

Wilson, S. T., Aylward, F. O., Ribalet, F., Barone, B., Casey, J. R., Connell, P. E., Eppley, J. M., Ferrón, S., Fitzsimmons, J. N., Hayes, C. T., Romano, A. E., Turk-Kubo, K. A., Vislova, A., Armbrust, E. V., Caron, D. A., Church, M. J., Zehr, J. P., Karl, D. M., and DeLong, E. F.: Coordinated regulation of growth, activity and transcription in natural populations of the unicellular nitrogen-fixing cyanobacterium *Crocospaera*, *Nat. Microbiol.*, 2, 17118, doi:<https://doi.org/10.1038/nmicrobiol.2017.118>, 2017.