



## Database of diazotrophs in global ocean: abundance, biomass and nitrogen fixation rates

Y.-W. Luo<sup>1</sup>, S. C. Doney<sup>1</sup>, L. A. Anderson<sup>2</sup>, M. Benavides<sup>3</sup>, I. Berman-Frank<sup>4</sup>, A. Bode<sup>5</sup>, S. Bonnet<sup>6</sup>, K. H. Boström<sup>7</sup>, D. Böttjer<sup>8</sup>, D. G. Capone<sup>9</sup>, E. J. Carpenter<sup>10</sup>, Y. L. Chen<sup>11</sup>, M. J. Church<sup>8</sup>, J. E. Dore<sup>12</sup>, L. I. Falcón<sup>13</sup>, A. Fernández<sup>14</sup>, R. A. Foster<sup>15</sup>, K. Furuya<sup>16</sup>, F. Gómez<sup>17</sup>, K. Gundersen<sup>18</sup>, A. M. Hynes<sup>19,\*</sup>, D. M. Karl<sup>8</sup>, S. Kitajima<sup>16</sup>, R. J. Langlois<sup>20</sup>, J. LaRoche<sup>20</sup>, R. M. Letelier<sup>21</sup>, E. Marañón<sup>14</sup>, D. J. McGillicuddy Jr.<sup>2</sup>, P. H. Moisander<sup>22,\*\*</sup>, C. M. Moore<sup>23</sup>, B. Mourinho-Carballido<sup>14</sup>, M. R. Mulholland<sup>24</sup>, J. A. Needoba<sup>25</sup>, K. M. Orcutt<sup>18</sup>, A. J. Poulton<sup>26</sup>, E. Rahav<sup>4</sup>, P. Raimbault<sup>6</sup>, A. P. Rees<sup>27</sup>, L. Riemann<sup>28</sup>, T. Shiozaki<sup>16</sup>, A. Subramaniam<sup>29</sup>, T. Tyrrell<sup>23</sup>, K. A. Turk-Kubo<sup>22</sup>, M. Varela<sup>5</sup>, T. A. Villareal<sup>30</sup>, E. A. Webb<sup>9</sup>, A. E. White<sup>21</sup>, J. Wu<sup>31</sup>, and J. P. Zehr<sup>22</sup>

<sup>1</sup>Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

<sup>2</sup>Department of Applied Ocean Science and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

<sup>3</sup>Instituto de Oceanografía y Cambio Global, Universidad de Las Palmas de Gran Canaria, 35017, Las Palmas de Gran Canaria, Spain

<sup>4</sup>The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan 52900, Israel

<sup>5</sup>Instituto Español de Oceanografía, Centro Oceanográfico de A Coruña, 15080 A Coruña, Spain

<sup>6</sup>IRD-INSU-CNRS, Laboratoire d'Océanographie Physique et Biogéochimique, UMR 6535, Centre d'Océanologie de Marseille, Aix Marseille Université, France

<sup>7</sup>Department of Natural Sciences, Linnaeus University, 39182 Kalmar, Sweden

<sup>8</sup>School of Ocean and Earth Science and Technology, University of Hawaii, Honolulu, Hawaii 96822, USA

<sup>9</sup>Department of Biological Sciences and Wrigley Institute for Environmental Studies, University of Southern California, Los Angeles, California 90089, USA

<sup>10</sup>Romberg Tiburon Center, San Francisco State University, Tiburon, California 94920, USA

<sup>11</sup>Department of Marine Biotechnology and Resources, National Sun Yat-sen University, Kaohsiung 80424, Taiwan

<sup>12</sup>Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, MT 59717, USA

<sup>13</sup>Laboratorio de Ecología Bacteriana, Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico

<sup>14</sup>Departamento de Ecología e Biología Animal, Universidade de Vigo, 36310 Vigo, Spain

<sup>15</sup>Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, 28359 Bremen, Germany

<sup>16</sup>Department of Aquatic Bioscience, University of Tokyo, Yayoi, Bunkyo, Tokyo, 113-8657, Japan

<sup>17</sup>Instituto Cavanilles de Biodiversidad y Biología Evolutiva, Universidad de Valencia, P.O. Box 22085, 46071 Valencia, Spain

<sup>18</sup>Department of Marine Science, University of Southern Mississippi, Stennis Space Center, Mississippi 39529, USA

<sup>19</sup>MIT-WHOI Joint Program in Oceanography/Applied Ocean Science and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

<sup>20</sup>Leibniz Institute for Marine Sciences, Duesternbrooker Weg 20, 24105 Kiel, Germany

<sup>21</sup>College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331, USA

<sup>22</sup>Ocean Sciences Department, University of California, Santa Cruz, California 95064, USA

<sup>23</sup>University of Southampton, National Oceanography Centre, Southampton, European Way, Southampton SO14 3ZH, UK

<sup>24</sup>Department of Ocean, Earth & Atmospheric Sciences, Old Dominion University, Norfolk, VA 23529, USA

<sup>25</sup>Division of Environmental and Biomolecular Systems, Oregon Health & Science University, Portland, Oregon 97239, USA

<sup>26</sup>National Oceanography Centre, Waterfront Campus, European Way, Southampton, SO14 3ZH, UK

<sup>27</sup>Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth PL1 3DH, UK

<sup>28</sup>Marine Biological Section, University of Copenhagen, 3000 Helsingør, Denmark

<sup>29</sup>Lamont Doherty Earth Observatory, Palisades, New York 10964, USA

<sup>30</sup>Marine Science Institute, University of Texas at Austin, Port Aransas, Texas 78382, USA

<sup>31</sup>Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL 33149, USA

\*currently at: Department of Marine Sciences, University of Georgia, Athens, GA 30602, USA

\*\*currently at: Department of Biology, University of Massachusetts Dartmouth, North Dartmouth, Massachusetts 02747, USA

Correspondence to: Y.-W. Luo (yluo@whoi.edu)

Received: 27 December 2011 – Published in Earth Syst. Sci. Data Discuss.: 13 February 2012

Revised: 25 July 2012 – Accepted: 14 August 2012 – Published: 31 August 2012

**Abstract.** Marine N<sub>2</sub> fixing microorganisms, termed diazotrophs, are a key functional group in marine pelagic ecosystems. The biological fixation of dinitrogen (N<sub>2</sub>) to bioavailable nitrogen provides an important new source of nitrogen for pelagic marine ecosystems and influences primary productivity and organic matter export to the deep ocean. As one of a series of efforts to collect biomass and rates specific to different phytoplankton functional groups, we have constructed a database on diazotrophic organisms in the global pelagic upper ocean by compiling about 12 000 direct field measurements of cyanobacterial diazotroph abundances (based on microscopic cell counts or qPCR assays targeting the *nifH* genes) and N<sub>2</sub> fixation rates. Biomass conversion factors are estimated based on cell sizes to convert abundance data to diazotrophic biomass. The database is limited spatially, lacking large regions of the ocean especially in the Indian Ocean. The data are approximately log-normal distributed, and large variances exist in most sub-databases with non-zero values differing 5 to 8 orders of magnitude. Reporting the geometric mean and the range of one geometric standard error below and above the geometric mean, the pelagic N<sub>2</sub> fixation rate in the global ocean is estimated to be 62 (52–73) Tg N yr<sup>-1</sup> and the pelagic diazotrophic biomass in the global ocean is estimated to be 2.1 (1.4–3.1) Tg C from cell counts and to 89 (43–150) Tg C from *nifH*-based abundances. Reporting the arithmetic mean and one standard error instead, these three global estimates are 140 ± 9.2 Tg N yr<sup>-1</sup>, 18 ± 1.8 Tg C and 590 ± 70 Tg C, respectively. Uncertainties related to biomass conversion factors can change the estimate of geometric mean pelagic diazotrophic biomass in the global ocean by about ± 70 %. It was recently established that the most commonly applied method used to measure N<sub>2</sub> fixation has underestimated the true rates. As a result, one can expect that future rate measurements will shift the mean N<sub>2</sub> fixation rate upward and may result in significantly higher estimates for the global N<sub>2</sub> fixation. The evolving database can nevertheless be used to study spatial and temporal distributions

and variations of marine N<sub>2</sub> fixation, to validate geochemical estimates and to parameterize and validate biogeochemical models, keeping in mind that future rate measurements may rise in the future. The database is stored in PANGAEA (doi:10.1594/PANGAEA.774851).

## 1 Introduction

N<sub>2</sub> fixation is the biological conversion of dinitrogen (N<sub>2</sub>) gas into two molecules of ammonia by diazotrophic organisms. Over geological time scales, N<sub>2</sub> fixation is important for regulating fixed N concentrations in the ocean and thereby sustaining ocean fertility (Tyrrell, 1999). The rate of pelagic N<sub>2</sub> fixation in the contemporary ocean has been estimated to be 100–200 Tg nitrogen (N) yr<sup>-1</sup>, which constitutes about half of the total external source of bioavailable N to the ocean (Gruber and Sarmiento, 1997, 2002; Karl et al., 2002; Galloway et al., 2004; Deutsch et al., 2007; Gruber, 2008). It is generally accepted that cyanobacteria are the major N<sub>2</sub>-fixing microorganisms in the ocean (Karl et al., 2002; Zehr, 2011). However, non-cyanobacterial prokaryotic plankton may also conduct N<sub>2</sub> fixation in the ocean as revealed by the presence and transcription of *nifH* genes (encoding the iron protein component of the nitrogenase enzyme) (Zehr et al., 1998; Riemann et al., 2010; Farnelid et al., 2011; Fernández et al., 2011), albeit their relative contribution to global N<sub>2</sub> fixation remains to be determined.

There are three major types of marine planktonic diazotrophic cyanobacteria (Sohm et al., 2011b; Zehr, 2011): (1) non-heterocystous filamentous cyanobacteria; (2) heterocystous cyanobacteria; and (3) unicellular cyanobacteria. Among them, most field and laboratory research has focused on non-heterocystous filamentous *Trichodesmium* (*Oscillatoriales*) (Dugdale et al., 1961; Carpenter and Romans, 1991; Capone et al., 1997). *Trichodesmium* is an abundant diazotroph in the open ocean (Capone et al., 2005; LaRoche and Breitbarth, 2005), which can be found as large colonies

known as “puffs” or “tufts” or as free trichomes (Carpenter, 1983).

Heterocystous cyanobacteria (*Nostocales*, of the genera *Richelia* and *Calothrix*) are frequently found in several oceanic diatom genera, including *Rhizosolenia* and *Hemiaulus* where cyanobacteria live within the diatom frustule but outside the cell wall, or as epiphytes on *Chaetoceros* diatoms (Mague et al., 1974, 1977; Venrick, 1974; Villareal, 1990; Foster and O’Mullan, 2008).

The most recently characterized diazotrophic phylotypes are the unicellular cyanobacteria (UCYN). Zehr et al. (2001) first successfully amplified a fragment of the *nifH* gene and *nifH* transcripts from the < 10 µm size fraction of the whole water samples that demonstrated the presence of unicellular diazotrophs. Subsequently, Montoya et al. (2004) measured high rates of N<sub>2</sub> fixation by UCYN in the Pacific Ocean. Three distinct phylogenetic groups have been identified from UCYN, including *Crocospaera watsonii* (sometimes referred to as Group B, UCYN-B), uncultivated Group A (UCYN-A) (Zehr et al., 2001), and Group C (UCYN-C) (Langlois et al., 2005; Foster et al., 2007) which have only recently been cultured (Taniuchi et al., 2012).

Although marine cyanobacterial diazotrophs play a critical role in the oceanic N cycle, primary productivity and organic matter export (e.g., Karl et al., 1997, 2002; Capone, 2000; Gruber, 2008), there is still no database synthesizing the many field measurements of diazotrophic abundances and N<sub>2</sub> fixation rates in the global ocean. Such a database is fundamental to understand the spatial distribution and temporal variability of diazotrophic biomass and activity. Moreover, a more comprehensive set of direct measurements may be useful in evaluating basin- and global-scale geochemical estimates of diazotrophic N inputs, which have generally found the global N budget to be in deficit, with total N sources significantly lower than the N sinks (Gruber, 2008). The database can also be expected to provide useful information with which to investigate the controlling mechanisms for marine diazotrophic distribution and activities.

In this paper we present a database compiling data on the abundance, biomass and N<sub>2</sub> fixation rate of diazotrophs in the global ocean. This effort is part of the Marine Ecosystem Model Intercomparison Project (MAREMIP), in which field measurement-based databases are constructed for biomass and related process rates for phytoplankton functional types (PFTs) (Buitenhuis et al., 2012). The databases, named “MARine Ecosystem DATA” (MAREDAT), include nine PFTs: diatoms, *Phaeocystis*, coccolithophores, diazotrophs, picophytoplankton, bacterioplankton, mesozooplankton, macrozooplankton and pteropods, and also a database for dissolved organic carbon. In addition to the database for diazotrophs presented in this paper, other databases are presented in other papers of this special volume. The MAREDAT databases will be used for the future model intercomparison studies in the MAREMIP projects, and are also available for public use.

In Sect. 2, the database is described, including information about database construction, data measurement methods, data quality control and conversion from diazotroph abundances to carbon (C) biomass. In Sect. 3, we present and discuss synthesized results from the database, including (1) the results of quality control, (2) spatial and temporal distribution of cell counts of diazotrophs, N<sub>2</sub> fixation rates and *nifH*-based abundances, (3) general characteristics of the datasets including mean N<sub>2</sub> fixation rates and diazotrophic C biomass as a function of geographical location and depth, (4) estimates for global N<sub>2</sub> fixation rates and diazotrophic C biomass and (5) limitations of the database and the global estimates. In Sect. 4, we draw conclusions and provide recommendations for appropriate use of this database.

## 2 Data and methods

The database is available at PANGAEA (doi:10.1594/PANGAEA.774851).

### 2.1 Database construction

#### 2.1.1 Database summary

Data comprised of three types of direct measurements, cell counts of diazotrophs, N<sub>2</sub> fixation rates, and *nifH*-based abundances via quantitative polymerase chain reaction (qPCR) assays, were compiled from the scientific literature and personal communication with researchers working in the field. The database contains a total of 12 151 data points including three sub-databases: (1) cell counts of diazotrophs with 5326 data points (Table 1a), (2) N<sub>2</sub> fixation rates with 3624 data points (Table 1b) and (3) diazotrophic abundances estimated from *nifH* copy abundances (referred to as *nifH*-based abundances hereafter) with 3201 data points (Table 1c). Note that the counts for *Trichodesmium* were reported in number of colonies, trichomes or cells, yet we maintain usage of the term “cell counts” to distinguish count-based methods from *nifH*-based abundances. The diazotroph abundances based on cell counts and *nifH* genes were converted into C biomass using conversion factors (discussed below). In each sub-database, the data were grouped into three taxonomic types: *Trichodesmium*, UCYN and heterocystous cyanobacteria (Table 1). A separate grouping is maintained for those N<sub>2</sub> fixation rates measured from whole seawater samples (Table 1b).

Each volumetric data point is identified by its sampling date, geographic location (latitude and longitude) and depth. Depth-integrated values were calculated for those vertical profiles with measurements available at three or more depths. By doing this, the profiles were linearly interpolated from surface to bottom sampling depth, i.e., N<sub>2</sub> fixation rates and diazotrophic abundances are considered to be negligible below the bottom depth. These calculated depth-integrated data points were not counted in the total points in Table 1 as

**Table 1a.** Summary of data points for cell counts of diazotrophs, including volumetric measurements of *Trichodesmium*, unicellular cyanobacteria and heterocystous cyanobacteria and their depth-integrals.

Investigator	Cell counts for diazotrophs					Region	References
	<i>Trichodesmium</i>	Unicellular	Heterocystous	Depth Integral <sup>b</sup>	Total Points <sup>c</sup>		
Bar-Zeev			26		26	Mediterranean Sea	Bar-Zeev et al. (2008)
Benavides				29 <sup>b</sup>	29	Subtropical N. Atlantic	Benavides et al. (2011)
Borstad	521			73	521	Tropical W. Atlantic	Borstad (1978)
Brzezinski			7	1	7	Subtropical N. Pacific	Brzezinski et al. (1998)
Carpenter	377			44	377	Tropical/subtropical W. Atlantic	Carpenter and Price (1977)
Carpenter/Romans	15				15	Subtropical N. Atlantic	Carpenter and Romans (1991)
Capone/Carpenter/Subramaniam				94 <sup>b</sup>	94	Subtropical N. Pacific	Sohm et al. (2011a)
Capone/Carpenter/Subramaniam	402		176	101 <sup>b</sup> + 166	679	Tropical/subtropical N. Atlantic	Carpenter et al. (2004); Capone et al. (2005)
Capone	22				22	Arabian Sea	Capone et al. (1998)
Chang	85			61 <sup>b</sup>	146	East China Sea	Chang et al. (2000)
Chen				47 <sup>b</sup>	47	South China Sea	Chen et al. (2003)
Chen	20				20	Upstream Kuroshio and South China Sea	Chen et al. (2008)
Chen				35 <sup>b</sup>	35	Northwest Pacific (Kuroshio)	Chen et al. (2011)
Dong				36 <sup>b</sup>	36	South China Sea	Dong et al. (2008)
Dugdale	15				15	Subtropical N. Atlantic (Stn. S)	Dugdale et al. (1961)
Fernández/Mouriño-Carballido/Marañón	201				201	Cross N.-S. Atlantic	Fernández et al. (2010); Mouriño-Carballido et al. (2011); Fernández et al. (2012)
Furuya/Kitajima	30		30		60	Tropical/subtropical N. Pacific	Kitajima et al. (2009)
Goering	51			6	51	Tropical Atlantic	Goering et al. (1966)
Gómez			439	33	439	Tropical/subtropical N. Pacific	Gómez et al. (2005)
Gómez			83	12	83	Tropical/subtropical S. Pacific	unpublished data
Hegde	135				135	Indian Ocean	Hegde et al. (2008)
Hulburt	19				19	Subtropical N. Atlantic	Hulburt (1962)
Hulburt	8				8	Subtropical N. Atlantic	Hulburt (1966)
Hulburt	26				26	Caribbean Sea	Hulburt (1968)
Kromkamp	44			2	44	W. Indian Ocean (African coast)	Kromkamp et al. (1997)
Lugomela	98			12	98	W. Indian Ocean (African coast)	Lugomela et al. (2002)
Letelier/Karl	23				23	Subtropical N. Pacific (ALOHA)	Letelier and Karl (1996)
Marumo/Asaoka				26 <sup>b</sup>	26	Cross N.-S. Pacific	Marumo and Asaoka (1974)
McCarthy/Carpenter	65			13	65	Subtropical N. Atlantic	McCarthy and Carpenter (1979)
Moore	6				6	Tropical/subtropical Atlantic	Moore et al. (2009)
Orcutt/Gundersen	35			2	37	Subtropical N. Atlantic (BATS)	Orcutt et al. (2001)
Poulton	341		341		682	Madagascar Basin	Poulton et al. (2009)
Sato	9				9	Tropical Atlantic	Sato et al. (1966)
Scranton	7				7	Subtropical N. Atlantic	Scranton (1984)
Steven/Glombitza	69				69	Tropical Atlantic	Steven and Glombitza (1972)
Tyrrell	315			16	315	Cross N.-S. Atlantic (AMT1-10)	Tyrrell et al. (2003)
Villareal	242		242	62	484	Subtropical N. Pacific	Villareal et al. (2011)
Villareal			186	9 <sup>b</sup> + 25	195	Subtropical N. Pacific	Villareal et al. (unpublished data)
Villareal	20		30		50	Gulf of Mexico	Villareal et al. (unpublished data)
Whilte	52		52	22	104	Gulf of California	White et al. (unpublished data)
Wu	21			3	21	South China Sea	Wu et al. (2003)
Total	3274	0	1612	930	5326		

<sup>a</sup> Computed from vertical profiles unless marked for those reported by data providers as depth-integrals. <sup>b</sup> Data are reported by data providers as depth-integrated biomass.

<sup>c</sup> Do not include depth-integrated profiles unless they are reported by data providers as depth-integrated values (as marked by <sup>b</sup>).

**Table 1b.** Summary of data points for N<sub>2</sub> fixation rates, including volumetric measurements of *Trichodesmium*, unicellular cyanobacteria and heterocystous cyanobacteria and their depth-integrals.

Investigator	N <sub>2</sub> Fixation Rates							Region	References
	<i>Trichodesmium</i>	Uni-cellular	Hetero-cystous	Whole Seawater	Depth Integral <sup>d</sup>	Total Points <sup>f</sup>	Method		
Bar-Zeev			12			12	C <sub>2</sub> H <sub>2</sub> reduction	Mediterranean Sea	Bar-Zeev et al. (2008)
Benavides		53		60	32 <sup>e</sup>	145	C <sub>2</sub> H <sub>2</sub> reduction and <sup>15</sup> N <sub>2</sub> assimilation	Subtropical N. Atlantic	Benavides et al. (2011)
Berman-Frank/Rahav				15	1	15	<sup>15</sup> N <sub>2</sub> assimilation	Red Sea	Rahav et al. (2012b)
Berman-Frank/Rahav				45	6	45	<sup>15</sup> N <sub>2</sub> assimilation	Mediterranean Sea	Rahav et al. (2011); Yogev et al. (2011); Rahav et al. (2012a)
Bonnet		27		20	10	47	<sup>15</sup> N <sub>2</sub> assimilation	Tropical Pacific	Bonnet et al. (2009)
Bonnet		80			9	80	<sup>15</sup> N <sub>2</sub> assimilation	Mediterranean Sea	Bonnet et al. (2011)
Church/Karl				332	50	332	<sup>15</sup> N <sub>2</sub> assimilation	Subtropical N. Pacific (ALOHA)	Church et al. (2009)
Carpenter	157				39	157	C <sub>2</sub> H <sub>2</sub> reduction	Tropical/subtropical W. Atlantic	Carpenter and Price (1977)
Capone/Carpenter/Subramaniam				180	40 <sup>e</sup> + 33	220	C <sub>2</sub> H <sub>2</sub> reduction and <sup>15</sup> N <sub>2</sub> assimilation	Subtropical N. Pacific	Sohm et al. (2011a)
Capone/Carpenter/Subramaniam	410		185		119 <sup>e</sup> + 118	714	C <sub>2</sub> H <sub>2</sub> reduction	Tropical/subtropical N. Atlantic	Capone et al. (2005)
Church				7		7	<sup>15</sup> N <sub>2</sub> assimilation	Subtropical N. Pacific	Fong et al. (2008)
Dore		13		24	6	37	<sup>15</sup> N <sub>2</sub> assimilation	Subtropical N. Pacific (ALOHA)	Dore et al. (2002)
Falcón		24			6	24	<sup>15</sup> N <sub>2</sub> assimilation	Subtropical N. Pacific (ALOHA) and Subtropical N. Atlantic	Falcón et al. (2004)
Fernández/Mouriño-Carballido/Marañón				115	32	115	<sup>15</sup> N <sub>2</sub> assimilation	Cross N.-S. Atlantic	Fernández et al. (2010); Mouriño-Carballido et al. (2011); Fernández et al. (2012)
Foster				1		1	<sup>15</sup> N <sub>2</sub> assimilation	Subtropical N. Pacific	Needoba et al. (2007)
Furuya/Kitajima		24		29		53	C <sub>2</sub> H <sub>2</sub> reduction	Tropical/subtropical N. Pacific	Kitajima et al. (2009)
Furuya/Shiozaki				103	11	103	<sup>15</sup> N <sub>2</sub> assimilation	Tropical/subtropical Pacific	Shiozaki et al. (2010)
Gandhi				28	7	28	<sup>15</sup> N <sub>2</sub> assimilation	Indian Ocean	Gandhi et al. (2011)
Goering	40				4	40	<sup>15</sup> N <sub>2</sub> assimilation	Tropical Atlantic	Goering et al. (1966)
Gundersen				24	2	24	C <sub>2</sub> H <sub>2</sub> reduction	Subtropical N. Pacific	Gundersen et al. (1976)
Karl	12	12		12	9	36	<sup>15</sup> N <sub>2</sub> assimilation	Subtropical N. Pacific (ALOHA)	Grabowski et al. (2008)
Kromkamp	15					15	C <sub>2</sub> H <sub>2</sub> reduction	W. Indian Ocean (African coast)	Kromkamp et al. (1997)
Mague			8		1	8	C <sub>2</sub> H <sub>2</sub> reduction	Subtropical N. Pacific	Mague et al. (1974)
Mague				46	8	46	C <sub>2</sub> H <sub>2</sub> reduction	Subtropical N. Pacific	Mague et al. (1977)
McCarthy/Carpenter	24				5	24	C <sub>2</sub> H <sub>2</sub> reduction	Subtropical N. Atlantic	McCarthy and Carpenter (1979)
Montoya		16			1	16	<sup>15</sup> N <sub>2</sub> assimilation	Subtropical N. Pacific	Montoya et al. (2004)
Montoya		5				5	<sup>15</sup> N <sub>2</sub> assimilation	Arafura Sea (near Australia)	Montoya et al. (2004)
Moore				46	10	46	<sup>15</sup> N <sub>2</sub> assimilation	Tropical/subtropical Atlantic	Moore et al. (2009)
Mulholland					30 <sup>e</sup>	30	<sup>15</sup> N <sub>2</sub> assimilation	Subtropical NW. Atlantic	Mulholland et al. (2012)
Mulholland				181	26	181	<sup>15</sup> N <sub>2</sub> assimilation	Subtropical NW. Atlantic	Mulholland et al. (unpublished data)
Mulholland				185	36	185	<sup>15</sup> N <sub>2</sub> assimilation	Subtropical N. Atlantic	Mulholland et al. (unpublished data)
Orcutt/Gundersen	31					31	<sup>15</sup> N <sub>2</sub> assimilation	Subtropical N. Atlantic (BATS)	Orcutt et al. (2001)
Raimbault				158	24	158	<sup>15</sup> N <sub>2</sub> assimilation	Tropical/subtropical S. Pacific	Raimbault and Garcia (2008)
Raimbault				197	33	197	<sup>15</sup> N <sub>2</sub> assimilation	Subtropical S. Pacific	Garcia et al. (2007)
Raimbault				71	12	71	<sup>15</sup> N <sub>2</sub> assimilation	Mediterranean Sea	Sandroni et al. (2007)
Rees				5		5	<sup>15</sup> N <sub>2</sub> assimilation	Eastern Mediterranean Sea	Rees et al. (2006)
Rees				21	7	21	<sup>15</sup> N <sub>2</sub> assimilation	Subtropical N. Atlantic (FeeP)	unpublished data
Rees				70		70	<sup>15</sup> N <sub>2</sub> assimilation	Cross N.-S. Atlantic (AMT14&15)	unpublished data
Rees				2		2	<sup>15</sup> N <sub>2</sub> assimilation	Western English Channel	Rees et al. (2009)
Rees				57	5	57	<sup>15</sup> N <sub>2</sub> assimilation	Tropical E. Atlantic	Turk et al. (2011)
Turk/Zehr				11	2	11	<sup>15</sup> N <sub>2</sub> assimilation	Tropical Atlantic	Goebel et al. (2010)
White/Letelier		11		12	4	23	<sup>15</sup> N <sub>2</sub> assimilation	Subtropical N. Pacific	White et al. (unpublished data)
White/Letelier		10		43		53	<sup>15</sup> N <sub>2</sub> assimilation	Subtropical N. Pacific	Watkins-Brandt et al. (2011)
White				134	19	134	<sup>15</sup> N <sub>2</sub> assimilation	Gulf of CA and Tropical Pacific	White et al. (2011)
Total	689	275	205	2234	757	3624			

<sup>d</sup> Computed from vertical profiles unless marked for those reported by data providers as depth-integrals. <sup>e</sup> Data are reported by data providers as depth-integrated N<sub>2</sub> fixation rates.

<sup>f</sup> Do not include integrated profiles unless they are reported by data providers as depth-integrated values (as marked by <sup>e</sup>).

**Table 1c.** Summary of data points for *nifH*-based abundances from qPCR assays, including volumetric measurements of *Trichodesmium*, unicellular cyanobacteria and heterocystous cyanobacteria and their depth-integrals.

Investigator	<i>nifH</i> -based abundances					Region	References
	<i>Trichodesmium</i>	Unicellular	Heterocystous	Depth Integral	Total Points <sup>§</sup>		
Bonnet	12	23	12		47	Tropical/subtropical S. Pacific	Bonnet et al. (2008)
Boström/Riemann			58		58	Baltic Sea	Boström et al. (2007)
Church	5	16		3	21	Subtropical N. Pacific (ALOHA)	Church et al. (2005a)
Church	24	48	24	8	40	Subtropical N. Pacific (ALOHA)	Church et al. (2005b)
Church	8	16	8	50 <sup>h</sup>	82	Subtropical N. Pacific	Fong et al. (2008)
Foster/Zehr	66	198	66	50	330	Tropical Atlantic	Foster et al. (2007)
Foster/Needoba/Zehr		18		3	18	Subtropical N. Pacific	Needoba et al. (2007)
Foster/Zehr	6				6	Red Sea	Foster et al. (2009a)
Foster/Subramaniam/Zehr	19	40	18	10	95	Tropical E. Atlantic	Foster et al. (2009b)
Foster/Zehr	17	42	17	14	76	Gulf of California	White et al. (unpublished data)
Liu	25	49	14	15	88	South China Sea	Kong et al. (2011)
Langlois/LaRoche	140	420		80	560	Tropical/subtropical N. Atlantic	Langlois et al. (2008)
Mazard/Scanlan		24		4	24	Arabian Sea	Mazard et al. (2004)
Moisander/Zehr	54	105	36	18	195	South China Sea	Moisander et al. (2008)
Moisander/Zehr	178	367	110	81	655	Tropical/subtropical S. Pacific	Moisander et al. (2010)
Mulholland	22	51	37		110	Subtropical NW. Atlantic	Mulholland et al. (2012)
Orcutt/Gundersen		18			18	Gulf of Mexico	Orcutt et al. (unpublished data)
Rees/Turk	29	58	27		114	Tropical E. Atlantic	Turk et al. (2011)
Turk/Zehr	154	300	154	80	608	Tropical Atlantic	Goebel et al. (2010)
<b>Total</b>	<b>759</b>	<b>1793</b>	<b>599</b>	<b>416</b>	<b>3201</b>		

<sup>§</sup> Do not include integrated profiles unless they are reported by data providers as depth-integrated values (as marked by <sup>h</sup>).

<sup>h</sup> Data are reported by data providers as depth-integrated values.

they are derived values. An additional 709 data points of diazotroph cell counts and N<sub>2</sub> fixation rates that were originally reported as depth-integrated values were also included. These 709 depth-integrated data points are counted in the total in Table 1 as they are independent from other data points. Each depth-integrated data point is identified by its sampling date, geographic location and integral depth.

The database also provides total diazotrophic C biomass (from cell counts and *nifH*-based abundances) for each sample by summing values from the three defined diazotrophic types: *Trichodesmium*, unicellular and heterocystous cyanobacteria. Total N<sub>2</sub> fixation rates are also provided: when whole seawater N<sub>2</sub> fixation rates are available, they

are used as total N<sub>2</sub> fixation rates. Otherwise, the total N<sub>2</sub> fixation rates are calculated by summarizing values from the three defined diazotroph types. In many samples, measurements were not available for all the three defined diazotrophic types. Also, these three defined types do not represent the full diazotrophic community. Thus, the derived totals (via summation) can be considered as the lower limits of diazotrophic biomass and activity.

Accessory data (including temperature, salinity and concentrations of nitrate, phosphate, iron and chlorophyll) are also provided if available.

### 2.1.2 Cell counts

Cell counts for diazotrophs (Table 1a) were largely performed by standard light microscopy whilst a number of samples were counted using epifluorescence microscopy with blue or green excitation (Orcutt et al., 2001; Chen et al., 2003, 2008, 2011; Carpenter et al., 2004; Sohm et al., 2011a; Villareal et al., unpublished North Pacific data). The cell counts are limited to *Trichodesmium* and heterocystous cyanobacteria, but no cell counts are available for UCYN (Table 1a), as UCYN-B can only be directly identified by epifluorescence microscopy and UCYN-A have not been microscopically identified. UCYN-C has only recently been microscopically identified (Taniuchi et al., 2012).

Most counts for *Trichodesmium* were reported in number of colonies or trichomes per volume, and in a few datasets in cell densities. In order to use a unified biomass conversion factor for *Trichodesmium* (discussed later), all the *Trichodesmium* counts were converted to number of trichomes assuming commonly used conversion factors of 200 trichomes colony<sup>-1</sup> and 100 cells trichome<sup>-1</sup> (Carpenter, 1983; Letelier and Karl, 1996; LaRoche and Breitbarth, 2005; Benavides et al., 2011). An exception is for the dataset of Carpenter et al. (2004) where conversion factors were measured in selected vertical profiles in three cruises in the tropical North Atlantic with averages of 137 (71–267), 224 (89–411) and 148 (56–384) trichomes per colony, respectively. In this case, the measured conversion factors are used for this specific dataset. Notably, the assumed conversion factor of 200 trichomes colony<sup>-1</sup> is consistent with values reported in Carpenter et al. (2004).

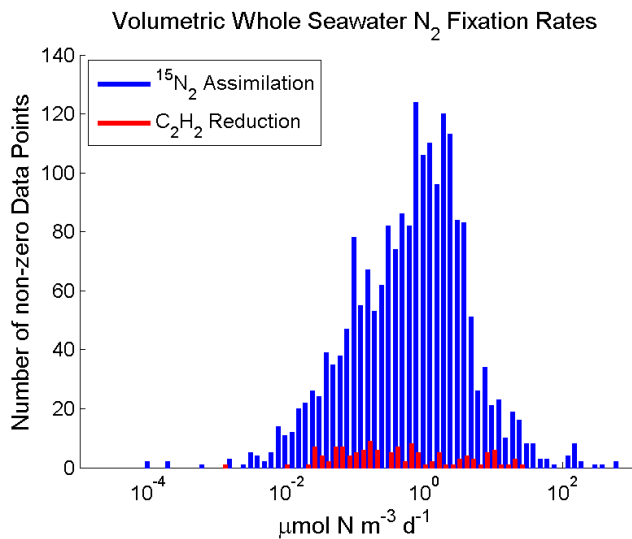
The cell counts for heterocystous bacteria were grouped into two major genera: *Richelia* and *Calothrix*. Counts for the *Richelia* and *Calothrix* are provided as heterocyst abundances. There are several datasets (Brzezinski et al., 1998; Gómez et al., 2005; Poulton et al., 2009; Villareal et al., 2011; Villareal et al., unpublished Gulf of Mexico data) in which abundances of host diatom *Hemiaulus* and *Rhizosolenia* were reported while *Richelia* heterocysts were not counted. As a *Hemiaulus* diatom typically contains 2 *Richelia* filaments whereas a *Rhizosolenia* diatom can contain 1–32 *Richelia* filaments (Sundström, 1984; Villareal, 1989, 1990; Foster and O'Mullan, 2008), the abundances of heterocysts for these datasets were derived from cell counts of their host diatoms by assuming that each *Hemiaulus* or *Rhizosolenia* cell contains 2 or 5 *Richelia* filaments, respectively. In one dataset (Gómez et al., 2005), abundances of *Chaetoceros* were counted but the associated heterocystous cyanobacteria were found to be *Richelia*. An average ratio of 0.5 *Richelia* heterocyst per *Chaetoceros* cell was reported by the dataset and was used to calculate the *Richelia* abundance. The trichomes or filaments of *Richelia* or *Calothrix* are typically composed of 3–4 vegetative cells and 1 terminal heterocyst (Foster and Zehr, 2006). Thus, abundances of cells within these genera are estimated by multiplying the heterocyst

abundances by 5, i.e., assuming 5 cells per filament. Note that it may underestimate heterocystous cell abundances as the *Richelia* symbionts of *Rhizosolenia* in some cases can contain more vegetative cells (near 10) (Villareal, 1989, 1992; Janson et al., 1999).

### 2.1.3 N<sub>2</sub> fixation rates

N<sub>2</sub> fixation rates were measured directly by <sup>15</sup>N-labeled N<sub>2</sub> gas (<sup>15</sup>N<sub>2</sub>) assimilation (Montoya et al., 1996) or indirectly by the acetylene (C<sub>2</sub>H<sub>2</sub>) reduction assay (Capone, 1993) (Table 1b). The <sup>15</sup>N<sub>2</sub> assimilation method tracks the conversion of <sup>15</sup>N<sub>2</sub> to particulate N. The <sup>15</sup>N<sub>2</sub> tracer is added into the ambient pool of N<sub>2</sub>; the <sup>15</sup>N/<sup>14</sup>N ratio is measured in the particulate N after incubation and compared to the natural abundance of N isotopes in unlabeled particulate material. The C<sub>2</sub>H<sub>2</sub> method estimates N<sub>2</sub> fixation rate indirectly by measuring the reduction of C<sub>2</sub>H<sub>2</sub> (a competitive inhibitor of N<sub>2</sub>) to ethylene (C<sub>2</sub>H<sub>4</sub>) which is then converted to a N<sub>2</sub> fixation rate assuming 3 or 4 moles C<sub>2</sub>H<sub>2</sub> reduced per 1 mole N<sub>2</sub> fixed, depending on the extent of nitrogenase-linked hydrogen production (Postgate, 1998). Generally, the direct <sup>15</sup>N<sub>2</sub> assimilation is a precise and sensitive method; hence, it has been used to generate the majority of rates (Table 1b). Direct comparison of these two methods showed the direct <sup>15</sup>N<sub>2</sub> assimilation method generally yields lower rates than those estimated from C<sub>2</sub>H<sub>2</sub> reduction assay (see summary in Mulholland, 2007). Discrepancies between these measures could be because the <sup>15</sup>N<sub>2</sub> assimilation method measures the net rate of conversion of reduced N to cellular N, or net N<sub>2</sub> fixation, while the C<sub>2</sub>H<sub>2</sub> reduction method measures gross N<sub>2</sub> fixation, which includes the reduced N both stored in cells and excreted as ammonium or dissolved organic N during incubation (Mulholland et al., 2004; Mulholland, 2007). More recently, it has been suggested that the direct <sup>15</sup>N<sub>2</sub> assimilation method significantly underestimates the N<sub>2</sub> fixation rates because the <sup>15</sup>N<sub>2</sub> bubbles injected in seawater do not attain equilibrium with surrounding water (Mohr et al., 2010), which will be discussed later. We include the N<sub>2</sub> fixation rates acquired by either the <sup>15</sup>N<sub>2</sub> assimilation method or the C<sub>2</sub>H<sub>2</sub> reduction assays in the database, considering that they distribute within a similar range of magnitude (Fig. 1). But this analysis does not consider different sampling sites. Further investigations using pair-wise comparison of the methods are needed to evaluate effects of these two methods on N<sub>2</sub> fixation rate measurement. Users have to be careful when using the database to study N<sub>2</sub> fixation rates aggregated from two different methods.

The collected N<sub>2</sub> fixation rates were mostly measured for whole seawater samples (Table 1b). Some samples were filtered and N<sub>2</sub> fixation rates were measured for organisms in the <10 μm size-fraction, which we have assigned to unicellular diazotrophs (Table 1b). Note that UCYN-B can form colonies and may not be included in this size fraction. It is also possible that some diatoms with associated



**Figure 1.** Histogram of volumetric, whole seawater  $N_2$  fixation rates (non-zero values) from  $^{15}N_2$  assimilation assays and  $C_2H_2$  reduction assays. Data values are on logarithmic scale.

heterocystous cyanobacteria may be included in  $< 10 \mu m$  fractions.  $N_2$  fixation rates were also measured in some datasets specifically for *Trichodesmium* and heterocystous cyanobacteria. Most  $N_2$  fixation data were reported as daily rates, except for 11 datasets that were reported as hourly rates. *Trichodesmium* fixes  $N_2$  exclusively during the light period, while the diel patterns of  $N_2$  fixation are unclear for other diazotrophs (Carpenter and Capone, 2008 and references therein). Thus, those hourly  $N_2$  fixation rates were converted to daily rates by multiplying by 12 h, which, however, could be conservative if diazotrophs other than *Trichodesmium* could fix  $N_2$  during night (e.g., Montoya et al., 2004; Zehr et al., 2007).

#### 2.1.4 *NifH*-based abundances

*NifH* abundances were estimated by qPCR targeting the *nifH* gene (Church et al., 2005a, b; Foster et al., 2007). Cellular DNA was extracted, and the gene sequences were targeted for different diazotrophic groups. Most *nifH*-based abundances were estimated for the three major diazotroph types: *Trichodesmium*, UCYN groups and heterocystous groups (Table 1c). Gene copies for UCYN were identified as UCYN-A, UCYN-B and/or UCYN-C groups. *NifH*-based abundances were also estimated for different groups of heterocystous cyanobacteria based on three *nifH* gene sequences (het-1, het-2 and het-3), which have been identified in symbioses with diatoms: *Richelia-Rhizosolenia*, *Richelia-Hemiaulus* and *Calothrix-Chaetoceros*, respectively (Church et al., 2005b; Foster and Zehr, 2006). Note that there is one dataset (Boström et al., 2007) that reports the abundance of heterocystous genus *Nodularia*. The sensitivity of the qPCR

thermocycler is usually in the range of 6–12 *nifH* copies per reaction. When accounting for sample filtration volume, final elution volume and the volume of the DNA extract used in the qPCR reaction, the detection limit of the assay is significantly higher. Due to the variety of factors that determine the assay detection limit, it varies between laboratories and even datasets and is typically on the order of 25–100 *nifH* copies  $l^{-1}$ . Thus, for those data points reported as “detected but not quantifiable”, 10 *nifH* copies  $l^{-1}$  could be a conservative estimate.

To estimate diazotrophic abundances, *nifH* gene copies are converted to number of diazotrophic cells on 1 : 1 basis, as it is for those diazotrophic genomes that have been sequenced (e.g., *Trichodesmium* and UCYN-B) that there is one *nifH* gene copy per genome (Zehr et al., 2008) and assuming one genome copy per cell. *NifH* genes are present in both the vegetative cells and the heterocysts of heterocystous cyanobacteria (Foster et al., 2009b). Thus, this estimate accounts for abundances of the total cells, not just the heterocysts, of the heterocystous cyanobacteria. Limitations are associated with this extrapolation. Evidence indicates that this extrapolation overestimates the diazotrophic abundances and can only be treated as an upper limit of the cell density because of the presence of multiple *nifH* gene copies per cell in some diazotrophs such as *Clostridium pasteurianum* (Langlois et al., 2008). Little information is available on the variability of genome copies per cell for all *nifH* phylotypes (Langlois et al., 2008). It is possible that, when there is excess phosphorus or if diazotrophic cells are carbon/energy limited rather than nutrient limited, they might accumulate more than one genome copy per cell. However, this extrapolation can also underestimate diazotrophic abundances as it is known that DNA and RNA extractions are not 100 % efficient and may vary among species (Foster et al., 2009b). The extraction efficiencies, and even genome copies per cell, are currently under investigation (J. Zehr, personal communication, 2011).

In addition, non-replicating deoxyribonucleic acid (DNA) can comprise up to 90 % of the total DNA in the oligotrophic regions (Winn and Karl, 1986) and DNA sampled in natural environments may represent non-living (detrital or non-replicating) particulate matter (Holm-Hansen et al., 1968; Holm-Hansen, 1969; Winn and Karl, 1986; Bailiff and Karl, 1991; Arin et al., 1999). There is also evidence suggesting that the non-living DNA is less important than originally thought (Dortch et al., 1983).

#### 2.2 Log-normal distribution and quality control

Diazotrophic abundances and  $N_2$  fixation rates in the ocean can range from 0 when diazotrophs are below detection or truly absent at that location and time, to very high values during diazotrophic bloom phases. Cell abundances and  $N_2$  fixation rates hence vary by several orders of magnitude, are often not normally distributed and are positively skewed (long tail of high values). However, the datasets (excluding



**Table 2.** Estimated default and the upper and lower bounds for the biomass conversion factors, and their impacts on global biomass estimates (based on geometric mean).

	Conversion Factor (pg C cell <sup>-1</sup> )			Global Biomass Estimate <sup>a</sup> (Tg C)		
	default	lower	upper	default	lower	upper
Cell-count-based biomass						
<i>Trichodesmium</i> <sup>b</sup>	300	100	500			
<i>Richelia</i>	10	2	80			
<i>Calothrix</i>	10	5	20			
				2.1	0.8	3.5
<i>nifH</i> -based biomass						
<i>Trichodesmium</i>	300	100	500			
UCYN-A	0.2	0.1	0.3			
UCYN-B	20	4	50			
UCYN-C	10	5	24			
UCYN all groups						
<i>Richelia</i>	10	2	80			
<i>Calothrix</i>	10	5	20			
				89	28	170

<sup>a</sup> The low/high biomass is calculated when all their subtypes use low/high conversion factors.

<sup>b</sup> Assuming 100 cells trichome<sup>-1</sup>.

zero-value data points) are approximately log-normally distributed (Figs. 2 and S1) as is typical of many biological and ecological properties that are induced by biological mechanisms (Koch, 1966; Campbell, 1995). Mathematically, calculating the (arithmetic) mean and standard error of the log-transformed data and then back-transforming them exponentially results in geometric mean  $\bar{x}_g$  and geometric standard error  $SE_g$ , which should be used in estimating the mean and the error of the mean for a log-normal distribution in format of  $\bar{x}_g \times SE_g$  ( $\bar{x}_g$  multiplied and divided by  $SE_g$ ) instead of using  $\bar{x}_a \pm SE_a$  ( $\bar{x}_a$ : arithmetic mean;  $SE_a$ : arithmetic standard error) (Limpert et al., 2001; Doney et al., 2009). Thus in this study, the error range for a geometric mean  $\bar{x}_g$  is represented as between  $\bar{x}_g/SE_g$  and  $\bar{x}_g \times SE_g$ .

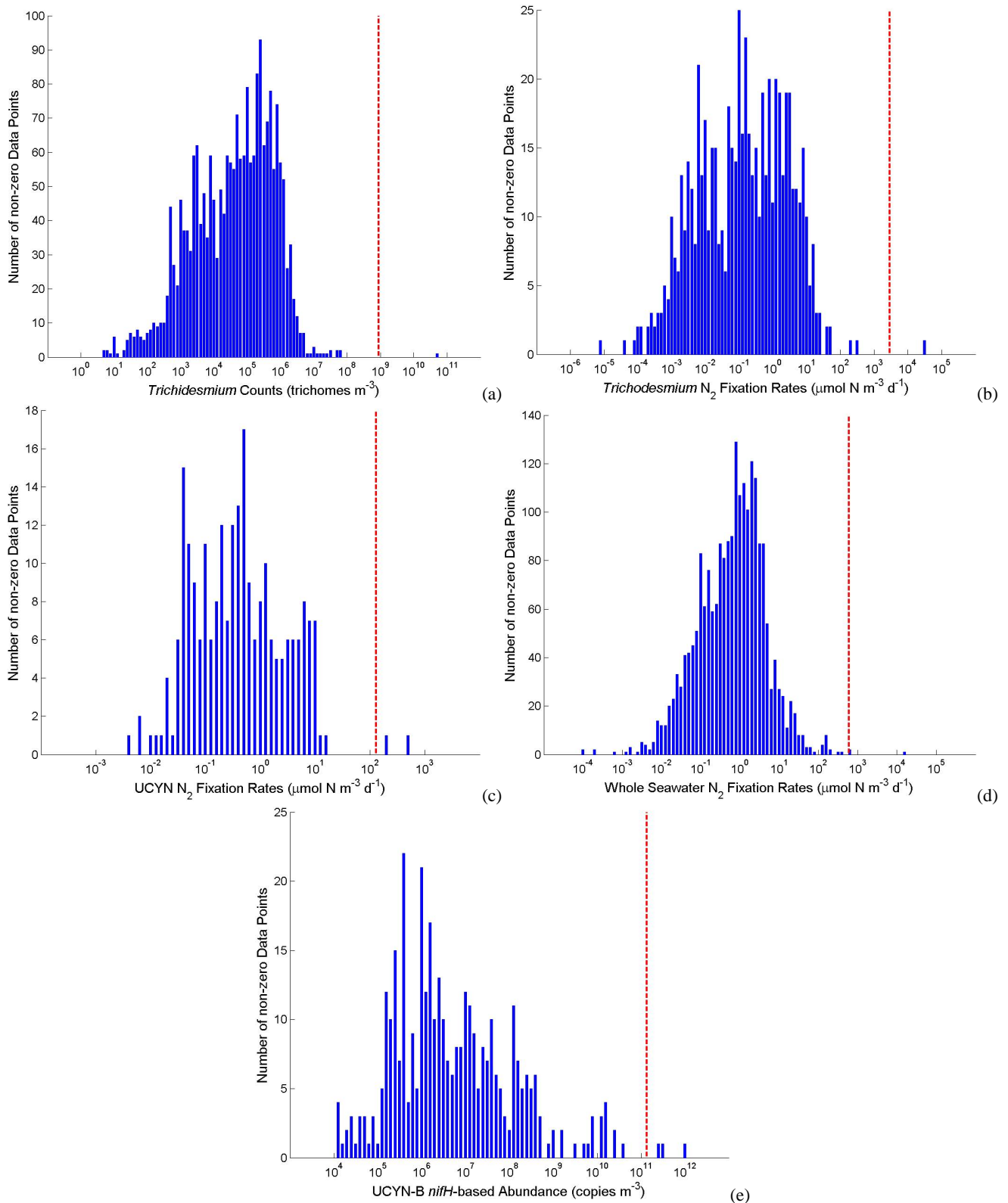
We control for data quality by using Chauvenet's criterion to flag suspicious outliers (Glover et al., 2011), which generally applies to normally distributed datasets and rejects data whose probability of deviation from the mean is less than  $1/(2n)$ , where  $n$  is the number of data points. However, considering (1) the datasets are log-normally distributed and (2) valid diazotrophic abundances and  $N_2$  fixation rates can be infinite low or zero, Chauvenet's criterion in our practice is applied to the log-transformed non-zero data to flag the outliers only on the high side. Those *nifH*-based abundances that were reported as "detected but not quantifiable" are not included in the application of Chauvenet's criterion. Although not used in the application of Chauvenet's criterion, data with zero values are kept in the database, as they represent valuable ecological information. Note that the criterion is processed separately for the volumetric data points of each taxonomic type. The criterion is also applied to the depth-

integrated total  $N_2$  fixation rates and the depth-integrated total biomass estimated from cell counts and *nifH*-based abundances, as these depth-integrated values are used later for the global estimates. First, the mean  $\bar{x}_{log}$  and the standard deviation  $\sigma_{log}$  of the log-transformed data are calculated, which are used to calculate the critical value  $x_{log}^*$  with a probability of a half of  $1/(2n)$  that values would exceed the mean  $\bar{x}_{log}$  by this amount assuming normal distribution (in log-transformed space). One-half of  $1/(2n)$  is used because Chauvenet's criterion is two-tailed test, and we only reject data at one tail on the high side. Thus, all data with log-transformed values higher than  $\bar{x}_{log} + x_{log}^*$  are flagged.

All the data points not flagged by the Chauvenet's criterion are accepted. However, not all the suspicious outliers flagged by the Chauvenet's criterion have to be rejected, as our datasets are not strictly log-normal distributed and the distribution estimated from the existing samples may not well represent the true distribution, especially at the sites with unusual environmental conditions. Thus for each flagged outlier, we evaluate whether its extraordinarily high value is reasonable or spurious based on the specific environmental conditions and/or the discussion with the original data contributor.

### 2.3 Biomass conversion

The cell counts and *nifH*-based abundances are converted to C biomass using conversion factors (Table 2). As discussed above, for C biomass estimates all the counts of *Trichodesmium* in colonies and cells have been converted to number of trichomes by assuming 200 trichomes colony<sup>-1</sup>



**Figure 2.** Histogram of data points on logarithmic scale (blue bars) and the critical values for quality control using Chauvenet's criterion (dashed red lines). Values higher than the critical values are rejected. (a) *Trichodesmium* cell counts, (b) *Trichodesmium*  $\text{N}_2$  fixation rates, (c) UCYN  $\text{N}_2$  fixation rates, (d) whole seawater  $\text{N}_2$  fixation rates, and (e) UCYN-B *nifH*-based abundance. See Supplement Fig. S1 for figures for other types.

**Table 3.** Biomass conversion factors estimated for *Trichodesmium* cells based on size measurements of cultured Woods Hole *Trichodesmium* species (Hynes et al., 2012). Carbon contents are calculated using the Verity et al. (1992) model.

Species	Biovolume ( $\mu\text{m}^3$ )		Carbon content ( $\text{pg C cell}^{-1}$ )	
	mean $\pm$ standard deviation	range	mean $\pm$ standard deviation	range
<i>T. thiebautii</i>	680 $\pm$ 380	280–1200	120 $\pm$ 57	55–190
<i>T. tenue</i>	610		110	
<i>T. pelagicum</i>	1100 $\pm$ 130	1000–1200	190 $\pm$ 18	170–200
<i>T. hildebrandtii</i>	1600 $\pm$ 140	1500–1700	250 $\pm$ 9	240–250
<i>T. erythraeum</i>	340 $\pm$ 200	190–920	65 $\pm$ 32	42–160
<i>T. contortum</i>	1300 $\pm$ 51	1300–1400	210 $\pm$ 7	210–220

and 100 cells trichome<sup>-1</sup>, which were widely used as an estimate for *Trichodesmium* (LaRoche and Breitbart, 2005). To determine the biomass conversion factor for trichomes of *Trichodesmium*, we utilize size measurements of cultured cells of different *Trichodesmium* species (Hynes et al., 2012) and estimate carbon content using the model of Verity et al. (1992) (Tables 3 and S1). The format of the Verity et al. (1992) model is

$$C = 0.433 \cdot V^{0.863}, \quad (1)$$

where  $C$  is the cell carbon content in  $\text{pg}$  and  $V$  is the cell volume in  $\mu\text{m}^3$ . This model was used because it was based on many data points across a wide range of cell sizes and species type, including cyanobacteria. *T. erythraeum* is the smallest *Trichodesmium* species, with estimated  $C$  content of  $65 \text{ pg C cell}^{-1}$  (Table 3) or  $6.5 \text{ ng C trichome}^{-1}$  (using 100 cells trichome<sup>-1</sup>), which is consistent with the back-calculated  $42 \pm 1 \text{ pg C cell}^{-1}$  from measurements of the iron-to-carbon ratio and iron content per cell (Tuit et al., 2004; Goebel et al., 2008). Thus, the Verity et al. (1992) model used here appears to be suitable for *Trichodesmium*. However, *T. erythraeum* is relatively small compared to other *Trichodesmium* species. The estimated carbon content for other *Trichodesmium* species is  $110\text{--}250 \text{ pg C cell}^{-1}$  (Table 3) or  $11\text{--}25 \text{ ng C trichome}^{-1}$  (using 100 cells trichome<sup>-1</sup>). Carbon content for *Trichodesmium* was derived from elemental analysis coupled with direct trichome counts at 40 stations in the tropical Atlantic in 1994 and 1996 (see the dataset associated with Carpenter et al., 2004 in our database), yielding a conversion factor of  $10 \pm 12 \text{ ng C trichome}^{-1}$ . Some other studies observed higher  $C$  contents for *Trichodesmium* colonies, such as  $9.7 \mu\text{g C colony}^{-1}$  in the Pacific (Mague et al., 1977) and  $10.9$  and  $11.6 \mu\text{g C colony}^{-1}$  for *Trichodesmium* puffs and tufts in Atlantic (McCarthy and Carpenter, 1979), which results in a conversion factor  $\sim 50 \text{ ng C trichome}^{-1}$  by assuming 200 trichomes colony<sup>-1</sup>. Another direct analysis of *Trichodesmium*  $C$  content at the BATS station in the Sargasso Sea in 1995–1997 shows  $8.3 \pm 3.5 \mu\text{g C colony}^{-1}$  ( $n = 23$ ) for *Trichodesmium* puffs,  $9.9 \pm 2.5 \mu\text{g C colony}^{-1}$  ( $n = 14$ ) for *Trichodesmium* tufts and  $4.5 \pm 2.4 \mu\text{g C colony}^{-1}$  ( $n = 6$ ) for

*T. erythraeum* (Orcutt and Gundersen, unpublished data), which, by assuming 200 trichomes colony<sup>-1</sup>, leads to a conversion factor of 41, 49 and  $22 \text{ ng C trichome}^{-1}$ , respectively. The synthesis of laboratory and field data by LaRoche and Breitbart (2005) provides values of  $50\text{--}58 \text{ ng C trichome}^{-1}$ . To accommodate all these estimates, we use default conversion factors of  $300 \text{ pg C cell}^{-1}$  ( $30 \text{ ng C trichome}^{-1}$ , using 100 cells trichome<sup>-1</sup>) for *Trichodesmium*, with its range estimated to  $100\text{--}500 \text{ pg C cell}^{-1}$  (Table 2).

The biomass conversion factor for UCYN-A is difficult to calculate because there is no isolate in culture. UCYN-A cells are spherical in shape and the estimate of  $\sim 1 \mu\text{m}$  in diameter by Goebel et al. (2008) is the only measurement for UCYN-A cell size, which was determined using fluorescence-activated cell sorting (FACS) coupled with real time-qPCR. This gives a cell size estimate of  $0.5 \mu\text{m}^3$  and a default conversion factor of  $0.2 \text{ pg C cell}^{-1}$  for UCYN-A using the Verity et al. (1992) model, with its range estimated by varying the default conversion factor by  $\pm 50\%$ , i.e.,  $0.1\text{--}0.3 \text{ pg C cell}^{-1}$  (Table 2).

UCYN-B (*Crocospaera*) cells are also spherical in shape with a reported diameter range of  $3\text{--}5 \mu\text{m}$  in laboratory isolates (Goebel et al., 2008) and  $3\text{--}8 \mu\text{m}$  in natural samples and cultures (Webb et al., 2009; Moisaner et al., 2010). Thus, UCYN-B cells have a range in volume from  $14\text{--}270 \mu\text{m}^3$  and cellular carbon content from  $4\text{--}50 \text{ pg C cell}^{-1}$  using the Verity et al. (1992) model (Table 2). By assuming a diameter of  $5 \mu\text{m}$  and thus a volume of  $65 \mu\text{m}^3$ , we also estimate the default conversion factor of  $20 \text{ pg C cell}^{-1}$  for UCYN-B.

The only successful isolation and laboratory culture of a UCYN-C strain, designated TW3 (Taniuchi et al., 2012), show that the cells are  $2.5\text{--}3.0 \mu\text{m}$  in width and  $4.0\text{--}6.0 \mu\text{m}$  in length, which gives cellular volume of  $20\text{--}42 \mu\text{m}^3$  and cellular carbon content of  $5\text{--}11 \text{ pg C cell}^{-1}$  for UCYN-C by using the Verity et al. (1992) model. However, this range is an estimate from only one UCYN-C strain. For example, the *nifH* gene of UCYN-C is most similar to the *nifH* gene of the benthic *Cyanothece* (Zehr, 2011), and the cell dimensions of two *Cyanothece* strains BH63 and BH68 have been reported as  $4\text{--}5 \mu\text{m}$  width by  $7\text{--}8 \mu\text{m}$  length (Reddy et al., 1993), which

**Table 4.** Biomass conversion factors estimated for *Richelia* and *Calothrix* trichomes based on biovolume measurements by Foster et al. (2011) and assuming trichome composition of one heterocyst and three, five or ten vegetative cells. Carbon contents are calculated using the Verity et al. (1992) model. Numbers are mean  $\pm$  standard deviation when applicable.

Species	Heterocyst		Vegetative cell		Vegetative cells per trichome	Biomass conversion (pg C cell <sup>-1</sup> )	
	Biovolume ( $\mu\text{m}^3$ )	C content (pg C cell <sup>-1</sup> )	Biovolume ( $\mu\text{m}^3$ )	C content (pg C cell <sup>-1</sup> )		mean	range
<i>Richelia</i> (n = 26)	110 $\pm$ 180	22 $\pm$ 31	42 $\pm$ 76	9 $\pm$ 14	3	13 $\pm$ 18	2–86
					5	12 $\pm$ 16	2–78
					10	11 $\pm$ 15	2–71
<i>Calothrix</i> (n = 5)	45 $\pm$ 35	11 $\pm$ 8	47 $\pm$ 19	12 $\pm$ 4	3	12 $\pm$ 5	7–18
					5	12 $\pm$ 5	7–17
					10	12 $\pm$ 4	8–17

equals 90–160  $\mu\text{m}^3$  and leads to a conversion factor of 15–24 pg C cell<sup>-1</sup> by using the Verity et al. (1992) model. By merging these two estimates, the default conversion factor of 10 pg C cell<sup>-1</sup> with a range of 5–24 pg C cell<sup>-1</sup> is used for UCYN-C (Table 2).

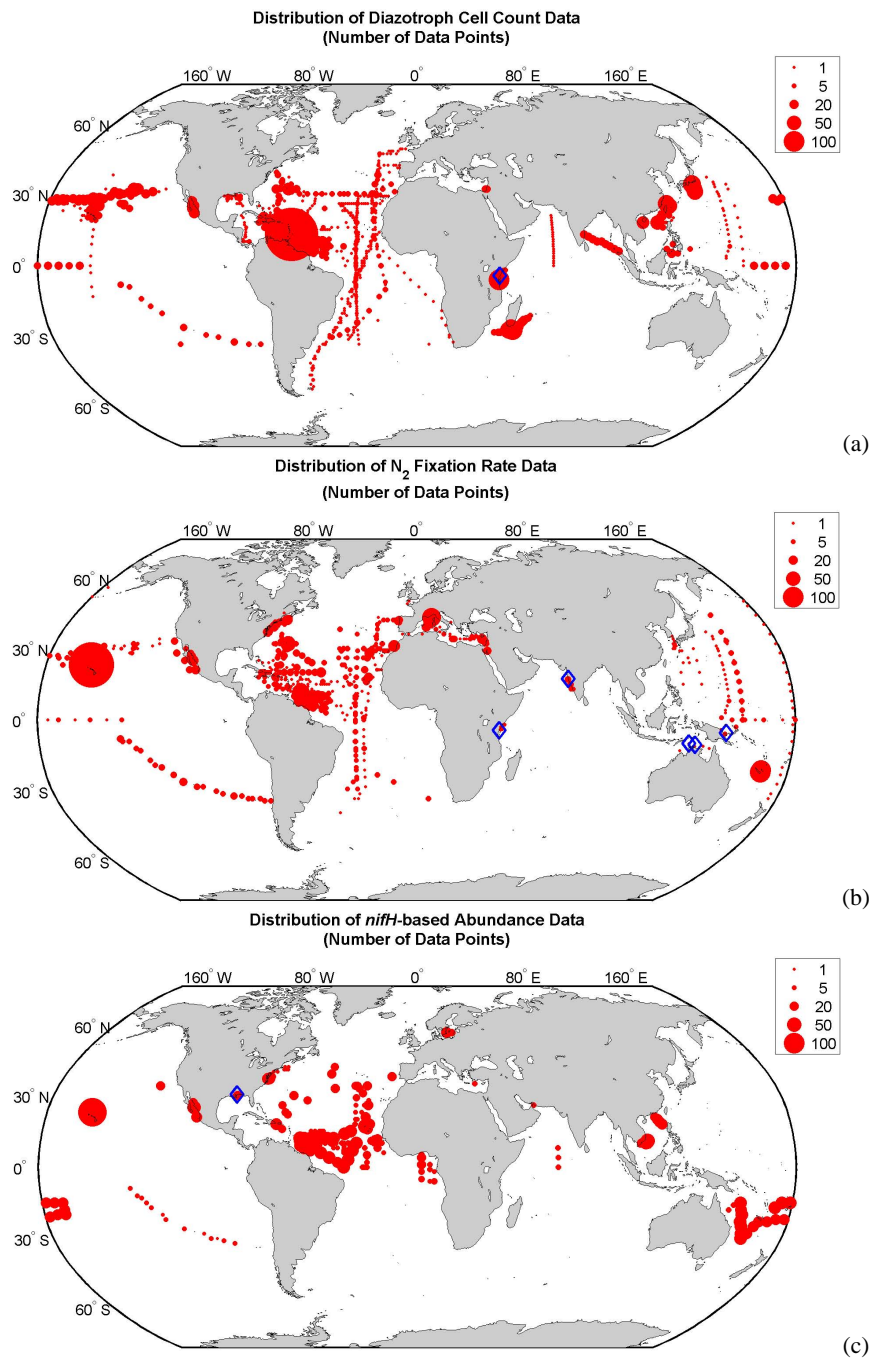
The trichomes of *Richelia* and *Calothrix* are comprised of three to ten vegetative cells and one terminal heterocyst (Janson et al., 1999; Foster and Zehr, 2006). The sizes of vegetative cells and heterocysts have been measured for multiple *Richelia* and *Calothrix* samples (Foster et al., 2011). We use these values to estimate the C contents of vegetative cells and heterocysts by using the Verity et al. (1992) model (Tables 4 and S2). As the sizes and the C contents are different in vegetative cells and heterocysts, average C content per *Richelia* or *Calothrix* cell is calculated by assuming each trichome is comprised of one heterocyst and three, five or ten vegetative cells (Table 4). The number of vegetative cells per trichome does not greatly impact the estimate of average C content per cell (Table 4). Based on these estimates, a default biomass conversion factor of 10 pg C cell<sup>-1</sup> with a range of 2–80 pg C cell<sup>-1</sup> is used for *Richelia*, and of 10 pg C cell<sup>-1</sup> with a range of 5–20 pg C cell<sup>-1</sup> for *Calothrix* (Table 2).

### 3 Results and discussion

#### 3.1 Results of quality control

Most data types are well approximated by a log-normal distribution (Figs. 2 and S1), except for *Calothrix* cell counts and *Calothrix nifH*-based abundances which only have limited non-zero data points (Fig. S1b and h). By applying Chauvenet's criterion, there are only 9 data points flagged, including 1 volumetric *Trichodesmium* cell count (Fig. 2a), 1 volumetric *Trichodesmium* N<sub>2</sub> fixation (Fig. 2b), 2 volumetric N<sub>2</sub> fixation rates by UCYN (Fig. 2c), 2 volumetric whole seawater N<sub>2</sub> fixation rate (Fig. 2d) and 3 volumetric UCYN-B *nifH* genes (Fig. 2e). Thus from a statistical point of view, most data are acceptable. The flagged 1 data point of *Trichodesmium* cell count of  $4.4 \times 10^{10}$  trichomes m<sup>-3</sup>

and 1 data point of *Trichodesmium* N<sub>2</sub> fixation rate of 31 391  $\mu\text{mol N m}^{-3} \text{d}^{-1}$  were sampled simultaneously in the western Indian Ocean near the Kenyan coast (Fig. 3a and b) (Kromkamp et al., 1997). The authors reported that a massive bloom, which seemed to be associated with a front, was encountered and large streaks with *Trichodesmium* colonies were floating at the surface. As both the high abundance and N<sub>2</sub> fixation rate were observed, these two data points were very likely real. The flagged data points of UCYN's N<sub>2</sub> fixation rate of 360 and 960  $\mu\text{mol N m}^{-3} \text{d}^{-1}$  (Montoya et al., 2004) were from  $\sim 10^\circ \text{S}$ ,  $\sim 130$ – $135^\circ \text{E}$ , the western tropical Pacific in the Arafura Sea (Fig. 3b). High N<sub>2</sub> fixation rates are commonly found in this region, such as another flagged whole seawater N<sub>2</sub> fixation rate of 610  $\mu\text{mol N m}^{-3} \text{d}^{-1}$  (Bonnet et al., 2009) found in the nearby sea region at  $6^\circ \text{S}$ ,  $147^\circ \text{E}$ , near Papua New Guinea (Fig. 3b). Recent measurements in the southwestern Pacific near New Caledonia also show extremely high N<sub>2</sub> fixation rates (Bonnet, unpublished data). The possible reason for these high N<sub>2</sub> fixation rates could be the iron supply from volcanoes or via upwelling. Another flagged whole seawater N<sub>2</sub> fixation rate of 13 500  $\mu\text{mol N m}^{-3} \text{d}^{-1}$  (Gandhi et al., 2011) was sampled from surface water at  $\sim 17^\circ \text{N}$ ,  $73^\circ \text{E}$ , near the Indian coast (Fig. 3b). In the same vertical profile, the N<sub>2</sub> fixation rates sampled in 2 m and deeper were all less than 32  $\mu\text{mol N m}^{-3} \text{d}^{-1}$ , and all the other surface N<sub>2</sub> fixation rates sampled in the surrounding area on this cruise were below 540  $\mu\text{mol N m}^{-3} \text{d}^{-1}$ . We evaluate that this high rate is extraordinary and should be removed from the database. The flagged three data points of UCYN *nifH*-based abundances of  $\sim 2$ – $9 \times 10^{11}$  copies m<sup>-3</sup> (Orcutt et al., unpublished data) are from the Mississippi Sound (Fig. 3c), although most other *nifH* gene data were sampled from the open ocean. Mississippi Sound is a very shallow, partially land-locked coastal environment and is very different from open ocean waters. Also, the three flagged data points of the extreme peaks in *nifH*-based abundance were sampled in June and July 2009 during extremely high water temperature events ( $\sim 30^\circ \text{C}$ ) towards the end of the summer. This is also a time of the year



**Figure 3.** Spatial distributions of collected diazotrophic data in number of data points (binned on to  $1^\circ \times 1^\circ$  grid), including cell counts (panel a),  $N_2$  fixation rates (panel b) and *nifH*-based abundances (panel c). Blue diamonds mark the location for the data rejected by Chauvenet's criterion.

when the lowest seasonal dissolved inorganic N:P-ratios (approximately 0.5) occurred, although the ratio increases towards 3–6 by the end of the late summer period (likely due to  $N_2$  fixation). The extreme peaks in *nifH*-based abundance also coincided with peaks in phytoplankton chlorophyll *a* ( $5\text{--}6 \mu\text{g l}^{-1}$ ). Thus, we believe most flagged data points, except the high surface  $N_2$  fixation rate found near the Indian coast

(Gandhi et al., 2011), are due to the specific environment and not necessarily related to the data quality. Hence, we have retained these values in the database. These points, however, are not included in our later analyses because their extremely high values would influence the mean values. It is also important to note that we have not excluded data based on an

assessment of the protocols of sampling, handling, preservation or measurement.

### 3.2 Data distribution

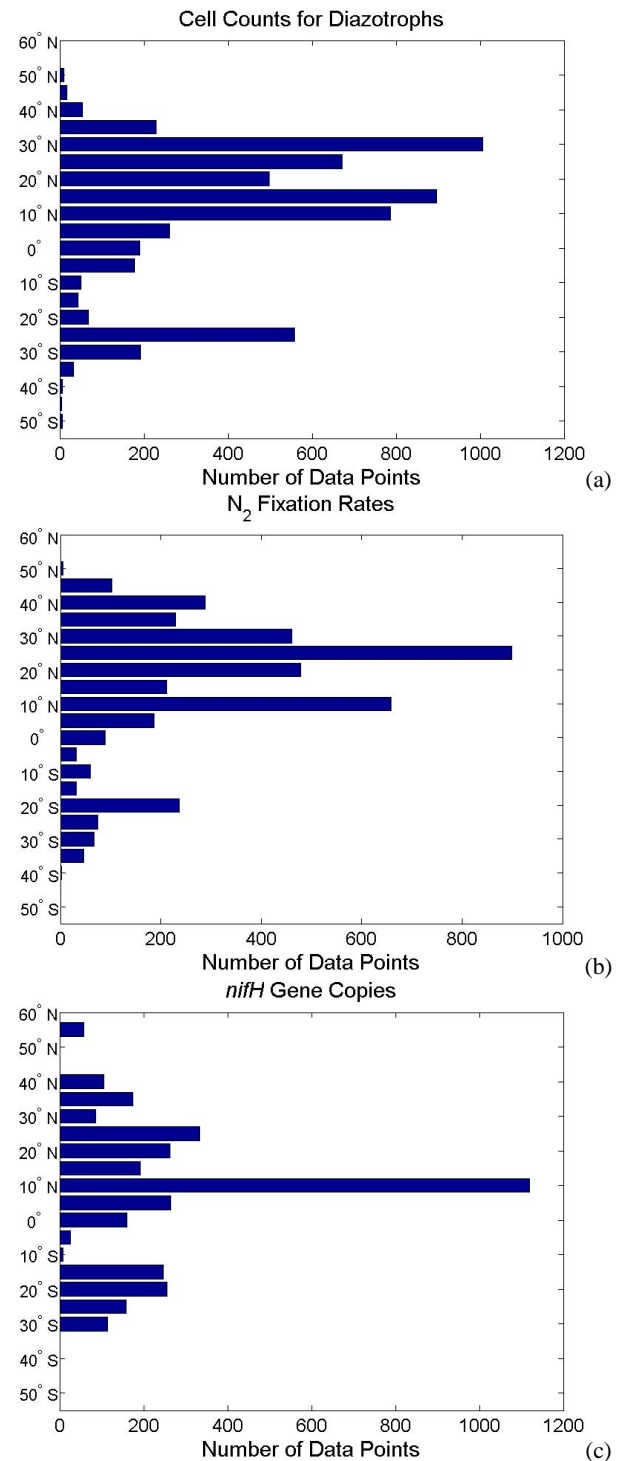
Figure 3 shows the spatial distribution of the three sub-databases. The Atlantic Ocean has the best data coverage in all three sub-databases, especially in the North Atlantic. In the Pacific Ocean, the coverage is limited especially in the South Pacific. There is almost no data coverage for the Indian Ocean except four datasets in the Arabian Sea (Capone et al., 1998; Mazard et al., 2004), near the Kenyan coast (Kromkamp et al., 1997) and in the Madagascar Basin (Poulton et al., 2009). There are also some data points in inner seas.  $N_2$  fixation rates were measured almost every month at the BATS station in 1995–1997 and at Station ALOHA from 2005–2010. These are the only two sites with long-term sustained time series of  $N_2$  fixation measurements.

Most data were collected in the tropical and subtropical regions, with latitudinal coverage of  $50^\circ\text{S}$ – $50^\circ\text{N}$  for cell counts of diazotrophs,  $40^\circ\text{S}$ – $60^\circ\text{N}$  for  $N_2$  fixation rates and  $30^\circ\text{S}$ – $60^\circ\text{N}$  for *nifH*-based abundances (Fig. 4). Most data were collected in the 1990s and 2000s, with some cell counts and  $N_2$  fixation rates collected in 1960s and 1970s and very limited data points of cell counts collected in 1980s (Fig. 5a, c and e). The monthly distribution of the data tends to be random for cell counts (Fig. 5b) and  $N_2$  fixation rates (Fig. 5d), while most of *nifH* gene data were collected in March, April and July (Fig. 5f).

### 3.3 $N_2$ fixation rates

After being binned onto  $3^\circ \times 3^\circ$  grids and calculating the geometric means of the data in each bin, depth-integrated  $N_2$  fixation rates are found to be highest in the western tropical Atlantic near the Caribbean Sea and in the subtropical North Pacific near the Hawaiian Islands, on an order of  $100$ – $1000 \mu\text{mol N m}^{-2} \text{d}^{-1}$  (Fig. 6a). In most other regions, depth-integrated  $N_2$  fixation rates are on an order of  $1$ – $100 \mu\text{mol N m}^{-2} \text{d}^{-1}$  (Fig. 6a).

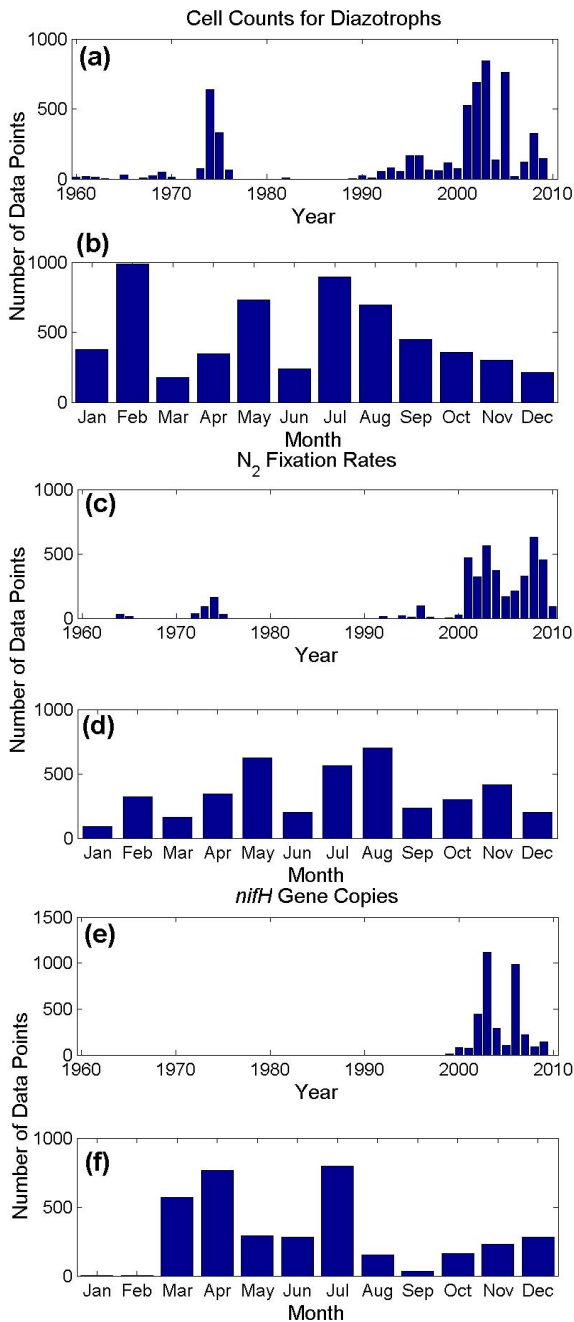
The volumetric  $N_2$  fixation rates are also analyzed on  $3^\circ \times 3^\circ$  grids in the five vertical layers of 0–5 m, 5–25 m, 25–62.5 m, 62.5–137.5 m and 137.5–250 m. The geometric mean values of each grid box are illustrated in Fig. 6b–f. Note that the depth-integrated and volumetric values are not from exactly the same data sources. They overlap for some data sources but also have locations that are not in common, as some of the data were originally reported as depth-integrated values and the other depth-integrated values were calculated herein.  $N_2$  fixation rates generally decrease with depth (Fig. 6b–f). When compared horizontally in each layer, the database reveals that there are several regions with high  $N_2$  fixation rates: the subtropical North Pacific in all the layers, surface waters of the western Pacific and the tropical Atlantic (especially in the west) in 0–25 m (Fig. 6b–f).



**Figure 4.** Latitudinal distribution of the data for (a) cell counts for diazotrophs, (b)  $N_2$  fixation rates, and (c) *nifH*-based abundances.

### 3.4 Diazotrophic abundance and biomass

Cell counts and *nifH*-based abundances are used to estimate diazotrophic C biomass using the default biomass conversion factors (Table 2), except for a few datasets where the



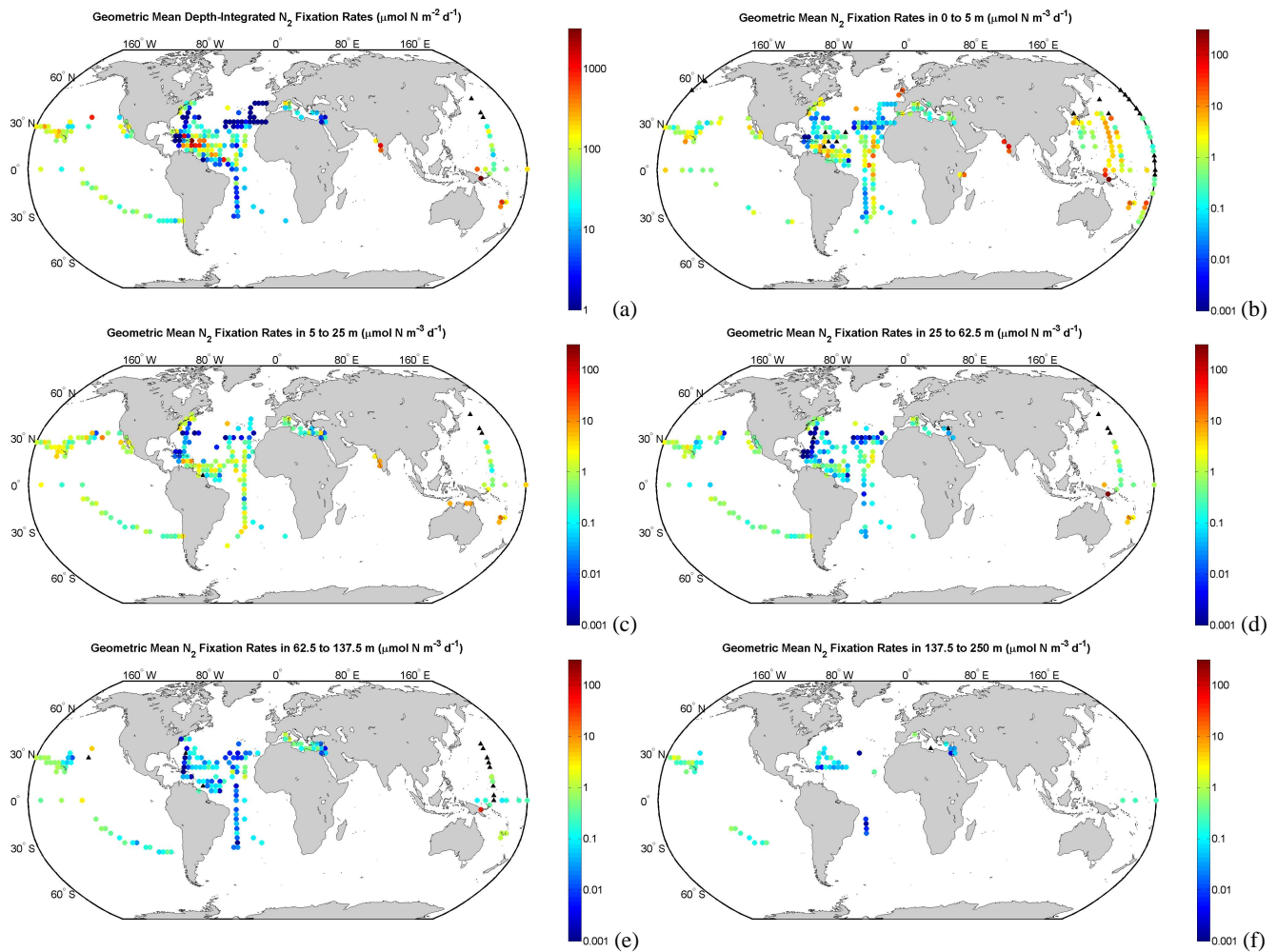
**Figure 5.** Temporal distribution of the data points in year and month for (a)–(b) cell counts for diazotrophs, (c)–(d)  $N_2$  fixation rates, and (e)–(f) *nifH*-based abundances.

contributors measured the conversion factors. The cell count data demonstrate that *Trichodesmium* is the dominant diazotroph, except that the abundance of *Calothrix* can be as large as *Trichodesmium* (Table 5). But the number of depth-integrated cell count samples for *Calothrix* are very limited (Table 5) and are limited to one cruise in the subtropical North Pacific (Villareal et al., unpublished). As the average *Trichodesmium* cell size, and thus the biomass con-

version factor, is much larger than heterocystous cyanobacteria (Table 2), *Trichodesmium* constitute more than 97% of diazotrophic biomass based on cell count data (Table 5). Cell count data do not include UCYN groups. The geometric mean of *nifH*-based abundances shows that UCYN-A is the most abundant among all the diazotrophic groups including *Trichodesmium*, and the abundances of UCYN-B, UCYN-C and *Richelia* are also comparable to *Trichodesmium* in the volumetric data (Table 5). However, *Trichodesmium* still dominate the diazotrophic biomass (Table 5) because of their higher biomass conversion factor than other groups. Comparison of the geometric means shows that *nifH*-based abundances are mostly one order of magnitude higher than cell-count-based abundances for *Trichodesmium* and heterocystous cyanobacteria, except that the volumetric *Trichodesmium* abundances are comparable in both cell-count-based and *nifH*-based geometric means (Table 5). In the North Atlantic Ocean where both cell counts and *nifH*-based abundances were frequently measured (Fig. 3a and c), histograms of *Trichodesmium* abundances derived from both the cell count data (assuming 100 cells trichome<sup>-1</sup>) and *nifH*-based data are in agreement (Fig. 7). This evidence provides some support for the reliability of *nifH*-based abundance for *Trichodesmium*.

For comparison, the arithmetic mean and standard error are also calculated for each diazotrophic group (Table 5). Note that zero-value data points have to be excluded for calculating geometric means, while these data points can be used to calculate arithmetic means. The arithmetic means are mostly one to several orders of magnitude larger than the geometric means, especially in the *nifH*-based datasets (Table 5). The estimated biomass of UCYN groups increases greatly and becomes comparable to that of *Trichodesmium* if the arithmetic means are used (Table 5), which, however, is certainly due to the high values within the approximate log-normal distributions dominating the calculation of the arithmetic mean.

Total diazotrophic C biomass estimated from cell counts is displayed in spatial maps for a given depth-integral and for the five vertical layers of 0–5 m, 5–25 m, 25–62.5 m, 62.5–137.5 m and 137.5–250 m by showing the geometric means of each 3°×3° grid bin (Fig. 8a–f). The depth-integrated cell-count-based diazotrophic biomass is higher in the western than in the eastern Atlantic (Fig. 8a). The depth-integrated cell-count-based diazotrophic biomass is also high in the subtropical North Pacific near Hawai’i, while it is low in other regions of the subtropical North Pacific because mostly only heterocystous cyanobacteria were counted in these sampling sites (Fig. 8a). The cell-count-based diazotrophic biomass shows maxima in the surface and decreases with depth (Fig. 8b–f). The surface cell-count-based diazotrophic biomass is high in the tropical Atlantic (Fig. 8b and c), which is consistent with the high  $N_2$  fixation rates found in the same region. In the southern Atlantic, the cell-count-based diazotrophic biomass is low in all layers (Fig. 8b–f). The high



**Figure 6.** (a) Geometric mean depth-integrated  $N_2$  fixation rates binned on  $3^\circ \times 3^\circ$  grid. Also geometric mean  $N_2$  fixation rates on  $3^\circ \times 3^\circ$  grid for 5 vertical layers of (b) 0–5 m, (c) 5–25 m, (d) 25–62.5 m, (e) 62.5–137.5 m and (f) 137.5–250 m. The color bar is in logarithmic scale. Zero values are marked with black triangles.

cell-count-based diazotrophic C biomass in the Arabian Sea is from one dataset (Capone et al., 1998) reporting an extensive *Trichodesmium* bloom, which may not represent mean level of diazotrophic biomass in the Arabian Sea.

The total diazotrophic biomass estimated from *nifH*-based abundances is presented for a given depth-integral and for the five vertical layers of 0–5 m, 5–25 m, 25–62.5 m, 62.5–137.5 m and 137.5–250 m by showing the geometric means of each  $3^\circ \times 3^\circ$  grid bin (Fig. 9a–f). Both the depth-integrals and results for vertical layers show high *nifH*-based C biomass in the tropical Atlantic (Fig. 9a–f), which is consistent with the high  $N_2$  fixation rates found in the same region. High *nifH*-based biomass was also found in the southwestern Pacific (Fig. 9a, c and d), where the diazotrophic cell count data are not reported. The *nifH*-based biomass is generally high in 0–62.5 m (Fig. 9b–d) and decreases below 62.5 m (Fig. 9e and f).

### 3.5 First-order estimates for global $N_2$ fixation rate and diazotrophic biomass

Many analyses can be conducted with  $\sim 12\,000$  data points in the database as per the objectives of the users. Here we show a simple example using the database to conduct first-order estimates of the global  $N_2$  fixation rate and diazotrophic biomass. We divide the global ocean into 6 regions: the North and the South Atlantic Ocean, the North and the South Pacific Ocean, the Indian Ocean and the Mediterranean Sea (Tables 6, 7 and 8). Two methods, geometric mean and arithmetic mean, were used for the estimation. As the data are not evenly distributed in space and intensive samplings were made in some regions (Fig. 3), the depth-integrated values are first binned to  $3^\circ \times 3^\circ$  grid to partially avoid this bias. Geometric and arithmetic means are calculated for each bin, which are then used to calculate geometric and arithmetic means for each region. Areal sum of the total  $N_2$  fixation



**Table 5.** Abundances and estimated biomass (using default conversion factors in Table 2) of each diazotrophic group from cell counts and *nifH*-based data, shown by geometric and arithmetic mean. Note that data points with zero-values are excluded when calculating geometric means. Error ranges for geometric means shown in parentheses are estimated by dividing and multiplying the geometric means by one geometric standard error. Standard errors for arithmetic means are shown in parentheses. Depth-integrated and volumetric data points are analyzed separately.

Depth-integrated Data	<i>n</i>	<i>n</i> of zero-value data	Abundance ( $10^9$ cells $m^{-2}$ )		Biomass ( $mg\ C\ m^{-2}$ )	
			Geometric	Arithmetic	Geometric	Arithmetic
cell-count-based data						
<i>Trichodesmium</i> *	620	59	0.16 (0.14–0.18)	1.2 (0.12)	43 (38–49)	310 (27)
<i>Richelia</i>	288	41	0.010 (0.008–0.013)	3.3 (0.88)	0.099 (0.078–0.13)	33 (8.8)
<i>Calothrix</i>	17	9	0.084 (0.050–0.14)	0.096 (0.050)	0.84 (0.50–1.4)	0.96 (0.50)
<i>nifH</i> -based data						
<i>Trichodesmium</i>	103	10	1.4 (1.1–1.9)	8.8 (1.5)	400 (310–530)	2600 (440)
UCYN-A	104	24	1.5 (0.88–2.4)	3400 (2200)	0.29 (0.18–0.48)	680 (450)
UCYN-B	103	20	0.11 (0.071–0.16)	3.1 (0.93)	2.1 (1.4–3.2)	63 (19)
UCYN-C	31	18	0.055 (0.020–0.16)	0.32 (0.16)	0.55 (0.20–1.6)	3.3 (1.6)
<i>Richelia</i>	63	2	0.24 (0.15–0.38)	34 (12)	2.4 (1.5–3.9)	340 (120)
<i>Calothrix</i>	11	0	0.21 (0.093–0.46)	0.58 (0.15)	2.1 (0.93–4.6)	5.8 (1.5)
Volumetric Data	<i>n</i>	<i>n</i> of zero-value data	Abundance ( $10^6$ cells $m^{-3}$ )		Biomass ( $\mu g\ C\ m^{-3}$ )	
			Geometric	Arithmetic	Geometric	Arithmetic
cell-count-based data						
<i>Trichodesmium</i> *	3222	1094	3.3 (3.1–3.5)	29 (3.6)	1100 (990–1100)	9400 (1100)
<i>Richelia</i>	1564	781	0.11 (0.097–0.12)	11 (1.5)	1.1 (0.97–1.2)	110 (15)
<i>Calothrix</i>	48	28	0.091 (0.060–0.14)	0.13 (0.043)	0.91 (0.60–1.4)	1.3 (0.43)
<i>nifH</i> -based data						
<i>Trichodesmium</i>	759	163	2.8 (2.5–3.2)	87 (12)	850 (740–970)	26 000 (3700)
UCYN-A	771	334	5.5 (4.5–6.8)	23 000 (9,400)	1.1 (0.89–1.4)	4600 (1900)
UCYN-B	792	322	1.1 (0.95–1.3)	280 (75)	22 (19–27)	5600 (1500)
UCYN-C	226	141	1.2 (0.90–1.7)	4.8 (1.1)	12 (9.0–17)	48 (11)
<i>Richelia</i>	535	160	1.6 (1.3–1.9)	240 (63)	16 (13–19)	2400 (630)
<i>Calothrix</i>	64	52	0.33 (0.17–0.62)	0.24 (0.10)	3.3 (1.7–6.2)	2.4 (1.0)

\* Assuming 100 cells trichome<sup>-1</sup>.

rate for each region is calculated by multiplying geometric or arithmetic means with ocean area of that region. Note that the volumetric data points are not used in this simple example, although they can also provide valuable information for the global estimates. Global  $N_2$  fixation rate and diazotrophic biomass are then estimated by summing the estimates from all the 6 regions. As 99% of the data in the database were collected within a latitudinal span of  $\sim 40^\circ S$ – $55^\circ N$ , we assume the oceanic  $N_2$  fixation is negligible outside of this latitudinal range. Given known temperature constraints on diazotrophy, this seems like a reasonable assumption; however, one could further refine global estimates if significant  $N_2$  fixation is found outside this latitudinal range in the future.

In the estimates for  $N_2$  fixation rate, geometric mean of the depth-integrated rates for the North Atlantic Ocean surprisingly is lower than that for the North and the South Pacific (Table 6), although the North Atlantic has attracted

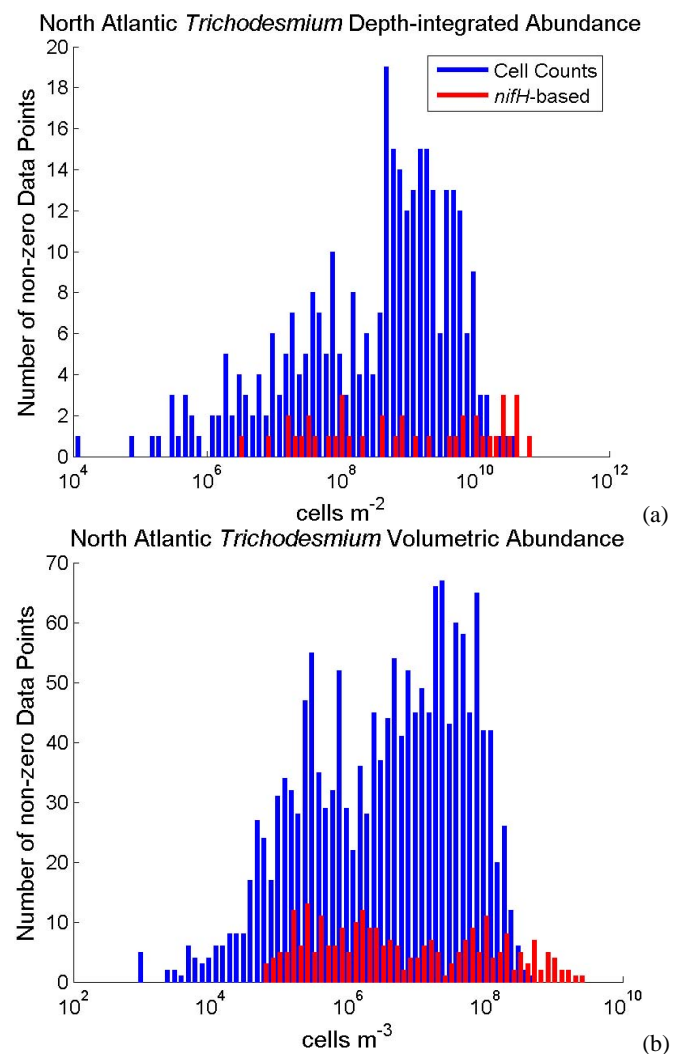
most studies of diazotrophs. By using the geometric means, the total  $N_2$  fixation rate in the North Atlantic is estimated to  $1.7\ Tg\ N\ yr^{-1}$ , one order lower than the estimates for the North and the South Pacific Ocean ( $35$  and  $24\ Tg\ N\ yr^{-1}$ , respectively) (Table 6). The arithmetic mean of  $N_2$  fixation rate in the North Atlantic, however, is one order higher than the geometric mean, which makes the total arithmetic rate of  $32\ Tg\ N\ yr^{-1}$  more comparable to the North Pacific ( $56\ Tg\ N\ yr^{-1}$ ) and the South Pacific ( $46\ Tg\ N\ yr^{-1}$ ) (Table 6). The data indicate that, although high  $N_2$  fixation rates were obtained in the North Atlantic, especially in the western tropical North Atlantic (Fig. 6a), low  $N_2$  fixation rates were more frequently found in this region. More measurements are needed to confirm the high estimates of  $N_2$  fixation rate for the Pacific Ocean, as the measurements in the North Pacific were mostly in the subtropical gyre near the Hawai'i Islands (Fig. 6a), and the sampling in the South Pacific has

been limited to data mostly from two cruises (Garcia et al., 2007; Raimbault and Garcia, 2008). The  $N_2$  fixation rates were also low in the South Atlantic (Table 6). There are inadequate depth-integrated  $N_2$  fixation rates collected in the Indian Ocean to estimate the means of this basin. By summing the geometric mean rates from all the regions, the global  $N_2$  fixation rate (not including the Indian Ocean) is estimated to be 62 (error range: 53–73)  $Tg N yr^{-1}$  (Table 6), which is at the low end of the current geochemical estimates of 100–200  $Tg N yr^{-1}$  for marine pelagic  $N_2$  fixation (Gruber and Sarmiento, 1997, 2002; Karl et al., 2002; Galloway et al., 2004; Deutsch et al., 2007; Gruber, 2008). Note that, as described in Sect. 2.2, the error range for a geometric mean in this study is represented as between the geometric mean multiplied/divided by the corresponding geometric standard error. The estimate of global  $N_2$  fixation rate using arithmetic mean is higher at 140 (standard error: 9.2)  $Tg N yr^{-1}$  (Table 6), which is within the range of the current geochemical estimates.

In a manner similar to our global estimates of the  $N_2$  fixation rate, we have also calculated global mean diazotrophic biomass from cell-count-based biomass data (Table 7). As in many sampling sites in the Pacific Ocean, only heterocystous cyanobacteria are counted (Table 1a) and thus very low total diazotrophic biomass was reported for these sites (Fig. 8a). We give special consideration for the Pacific by estimating mean biomass for *Trichodesmium* and heterocystous cyanobacteria separately before summing the values from these two taxonomies (Table 7). The average diazotrophic biomass in the North Atlantic is similar to or higher than that in the North Pacific as shown by both geometric and arithmetic means (Table 7). The global diazotrophic biomass (not including Mediterranean Sea) is estimated from the cell-count-based data as 2.1 (error range: 1.4–3.1)  $Tg C$  using geometric mean or 18 (standard error: 1.8)  $Tg C$  using arithmetic mean (Table 7).

The global diazotrophic biomass is also estimated from depth-integrated *nifH*-based data. First, the depth-integrated data points of *nifH*-based biomass are limited, with no or inadequate data points in the South Atlantic, the Indian Ocean and the Mediterranean Sea (Table 8), and thus the estimate could be highly biased. The geometric mean *nifH*-based diazotrophic biomass is extremely high in the South Pacific compared to other regions (Table 8). The global estimate (the Pacific plus the North Atlantic) of diazotrophic biomass from *nifH*-based data is 89 (error range: 52–150)  $Tg C$  using geometric mean or 590 (standard error: 70)  $Tg C$  (using arithmetic mean), which are dominated by the estimate of the South Pacific (Table 8).

By comparing the global estimate of  $N_2$  fixation rates with the global estimates of cell-count-based or *nifH*-based diazotrophic biomass (using geometric means) (Tables 6, 7 and 8) and assuming molar C : N ratio of 106 : 16 for diazotrophic cells, the turnover time of diazotrophic cellular N

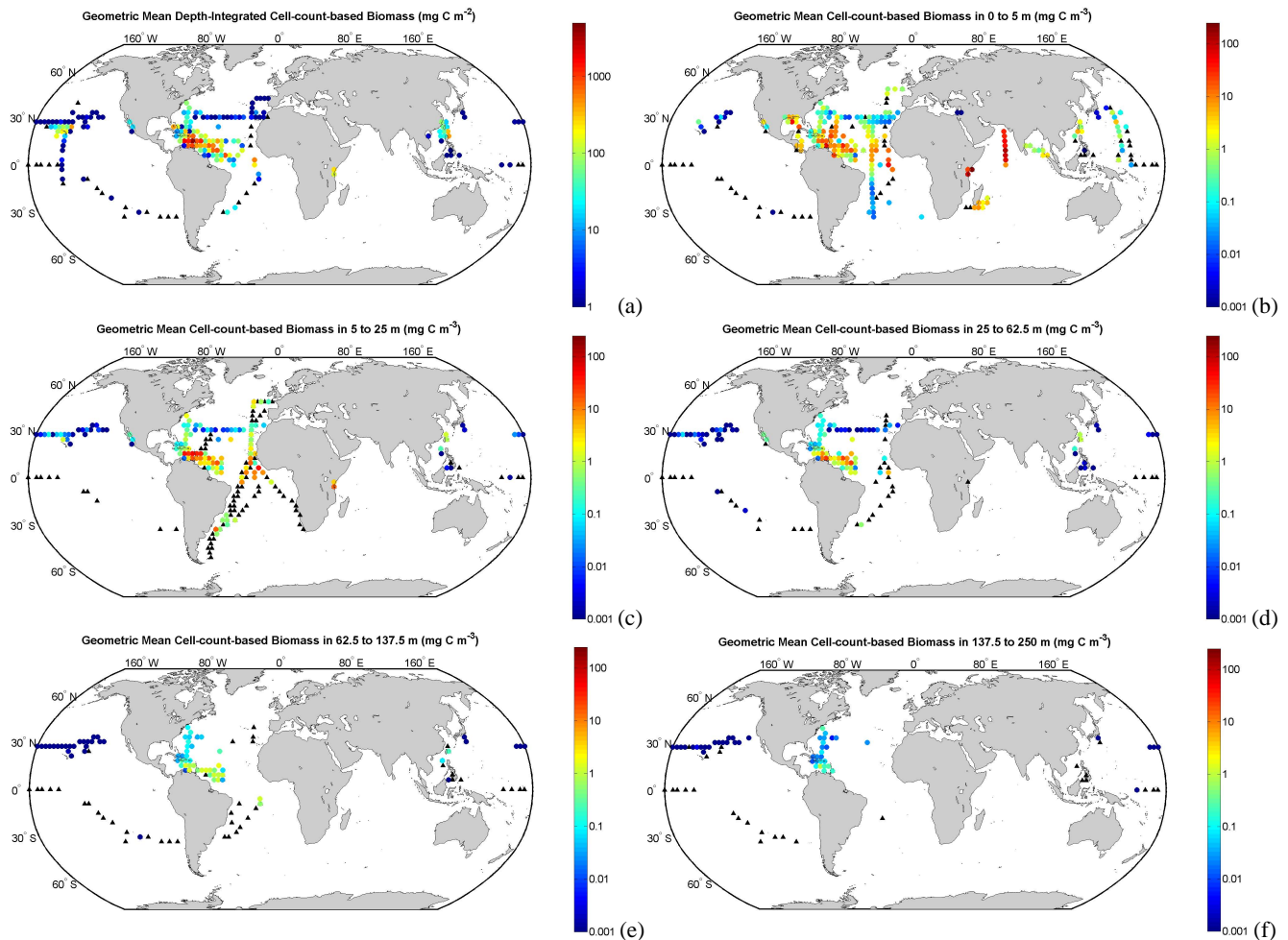


**Figure 7.** Histogram of depth-integrated (panel a) and volumetric (panel b) non-zero data points of cell-count-based (blue) and *nifH*-based (red) *Trichodesmium* abundances in the North Atlantic ( $10^{\circ} S$ – $50^{\circ} N$ ). Each *Trichodesmium* trichome is assumed to be comprised of 100 cells. Data values are on logarithmic scale. Those “detected but not quantifiable” *nifH*-based abundances, which are assigned  $10 \times 10^3$  cells  $m^{-3}$  (see text for details), are not included.

due to  $N_2$  fixation is estimated to be 2 days and 92 days, respectively.

### 3.6 Limitations

There are a number of limitations with the database and the global estimates. First and foremost, these data points are not uniformly distributed in the world ocean. Historically, scientists have also sought out regions with high diazotrophic abundances and, hence, there is a higher possibility of artificially elevated  $N_2$  fixation activity and diazotrophic biomass relative to true regional means. As the current database does not cover some regions such as the coastal upwelling zones,



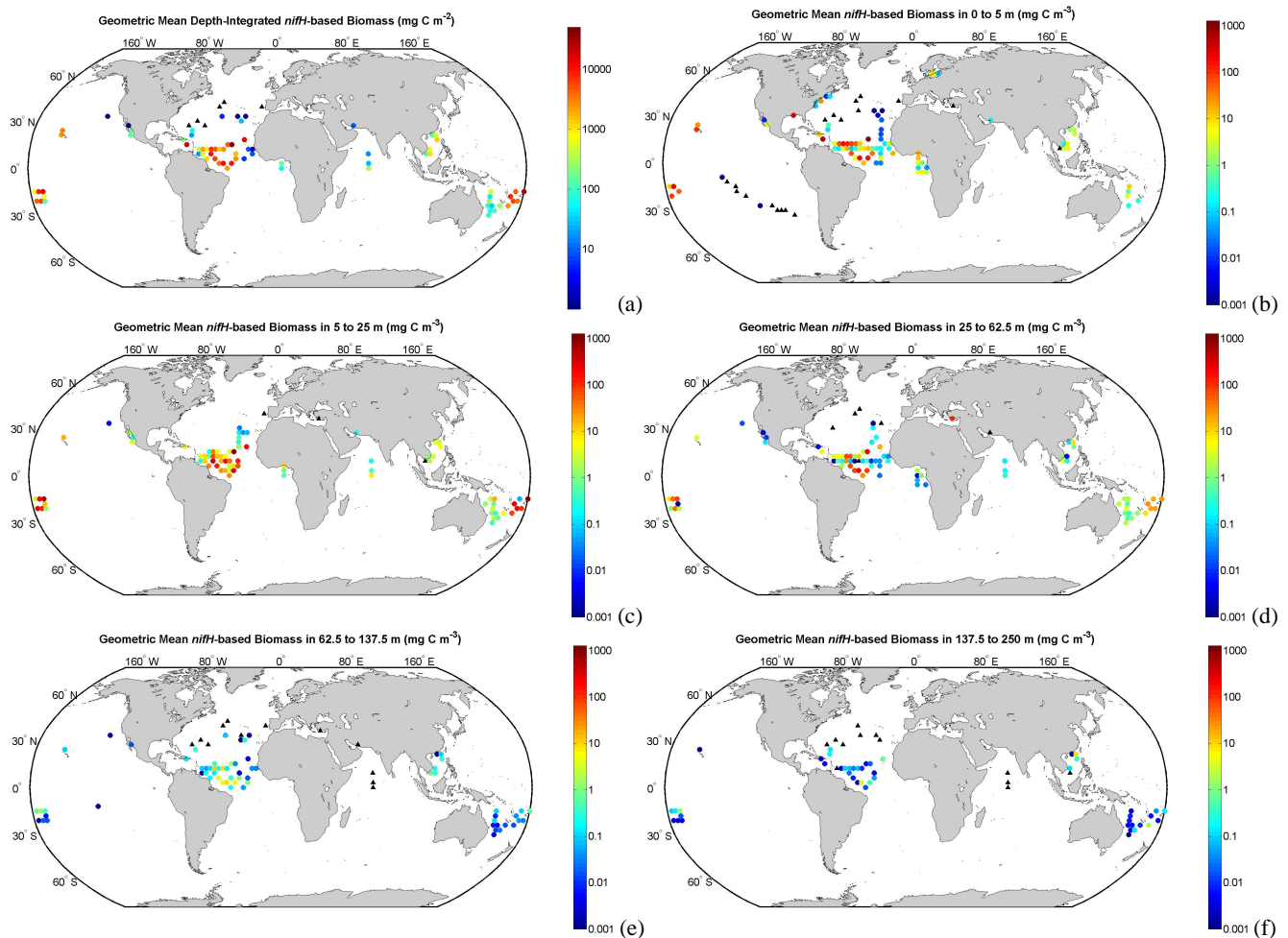
**Figure 8.** (a) Geometric mean depth-integrated cell-count-based diazotrophic carbon biomass binned on  $3^\circ \times 3^\circ$  grid. Also geometric mean cell-count-based diazotrophic carbon biomass on  $3^\circ \times 3^\circ$  grid for 5 vertical layers of (b) 0–5 m, (c) 5–25 m, (d) 25–62.5 m, (e) 62.5–137.5 m and (f) 137.5–250 m. The color bar is in logarithmic scale. Zero values are marked with black triangles.

our estimate could be changed substantially if values differ within these biogeographic domains.

As mentioned above, it was recently established that the most commonly applied method used to measure  $N_2$  fixation, the  $^{15}N_2$  assimilation method (Table 1b), has underestimated the true rates because the  $^{15}N_2$  bubbles injected in seawater do not attain equilibrium with surrounding water (Mohr et al., 2010). Two independent method comparisons conducted in the Pacific and Atlantic Oceans demonstrated that the measured rates are significantly higher when the  $^{15}N_2$  tracer is added in a dissolved form to the incubations, resulting in an instantaneous  $^{15}N$  isotopic equilibrium in the seawater (Großkopf et al., 2012; Wilson et al., 2012). The variability in the underestimation was linked to the composition of the diazotrophic community (Mohr et al., 2010; Großkopf et al., 2012) indicating that it will be difficult to correct the historical rate measurements made with the  $^{15}N_2$  method. However, these measurements can be treated as min-

imum estimates that may increase in the future. These measurements are still valuable when comparing  $N_2$  fixation rates in different regions, as they are approximately log-normally distributed and range in  $\sim 6$  orders of magnitude (Fig. 1).

It is also a notable issue that the difference between the geometric and the arithmetic mean  $N_2$  fixation rate is larger in the North Atlantic than those in the other basins (Table 6). A strict log-normal distribution is symmetric in the log space and its geometric mean is located at the peak of the distribution, i.e., at the highest probability. However, the depth-integrated total  $N_2$  fixation rates in our database, after binned on  $3^\circ \times 3^\circ$  grid, are not strictly log-normally distributed in each basin (Fig. 10). The North Atlantic has highest spatial coverage, and the  $N_2$  fixation rates in this basin differ in 6 orders of magnitude from  $10^{-3}$ – $10^3 \mu\text{mol N m}^{-2} \text{d}^{-1}$  (Fig. 10). Its distribution is not symmetric, with more values observed on the left (lower) side. Thus, the geometric mean  $N_2$  fixation rate for the North Atlantic is one order of magnitude lower



**Figure 9.** (a) Geometric mean depth-integrated *nifH*-based diazotrophic carbon biomass binned on  $3^\circ \times 3^\circ$  grid. Also geometric mean *nifH*-based diazotrophic carbon biomass on  $3^\circ \times 3^\circ$  grid for 5 vertical layers of (b) 0–5 m, (c) 5–25 m, (d) 25–62.5 m, (e) 62.5–137.5 m and (f) 137.5–250 m. The color bar is in logarithmic scale. Zero values are marked with black triangles.

**Table 6.** Estimates of  $N_2$  fixation rate for the global oceans based on geometric and arithmetic means. Note that data points with zero-values are excluded when calculating geometric means. Depth-integrated  $N_2$  fixation rates are first binned to  $3^\circ \times 3^\circ$  grid, and geometric and arithmetic means are calculated for each bin, which are then used to calculate geometric and arithmetic means for each region. Areal sum is calculated by multiplying geometric or arithmetic means with ocean area. Error ranges for geometric means shown in parentheses are estimated by dividing and multiplying the geometric means by one geometric standard error. Standard errors for arithmetic means are shown in parentheses. Error range for global geometric mean is estimated by simply summing lower and upper bounds of the error ranges of each region. Errors for global arithmetic means are propagated from (independent) errors in each region. ND: no data. NQ: not qualified to calculate areal sum because of too few samples.

Region	Latitudinal range	Ocean area ( $\times 10^{12} \text{ m}^2$ )	<i>n</i> of bins	<i>n</i> of zero-value bins	$N_2$ fixation rate ( $\mu\text{mol N m}^{-2} \text{ d}^{-1}$ )		Areal Sum ( $\text{Tg N yr}^{-1}$ )	
					Geometric	Arithmetic	Geometric	Arithmetic
North Atlantic Ocean	0–55° N	36	125	0	9.2 (7.0–12)	170 (40)	1.7 (1.3–2.2)	32 (7.4)
South Atlantic Ocean	40–0° S	27	15	0	7.9 (6.3–9.8)	13 (4.4)	1.1 (0.9–1.4)	1.8 (0.6)
North Pacific Ocean	0–55° N	89	45	3	78 (67–90)	120 (22)	35 (30–41)	56 (9.8)
South Pacific Ocean	40–0° S	72	26	0	64 (54–76)	130 (46)	24 (20–28)	46 (17)
Indian Ocean	40° S–25° N	62	4	0	120 (29–490)	590 (320)	NQ	NQ
Mediterranean Sea	30–45° N	2.5	10	0	18 (12–28)	45 (21)	0.2 (0.1–0.4)	0.6 (0.3)
Global							62 (52–73)	137 (9.2)

**Table 7.** Estimates of diazotrophic biomass for the global oceans using cell-count-based data. Same method as Table 6 is used, except that the estimates for *Trichodesmium* and Heterocystous cyanobacteria (CYN) are calculated separately in the North and South Pacific Ocean. ND: no data. NQ: not qualified to calculate areal sum because of too few samples.

Region	Latitudinal range	Ocean area ( $\times 10^{12}$ m <sup>2</sup> )	<i>n</i> of bins	<i>n</i> of zero-value bins	Cell-count-based diazotrophic biomass (mg C m <sup>-2</sup> )		Areal Sum (Tg C)	
					Geometric	Arithmetic	Geometric	Arithmetic
North Atlantic Ocean	0–55° N	36	103	4	24 (18–32)	310 (66)	0.87 (0.67–1.2)	11 (2.4)
South Atlantic Ocean	40–0° S	27	8	4	11 (6.7–18)	7.5 (3.9)	0.30 (0.18–0.50)	0.21 (0.11)
North Pacific Ocean	0–55° N	89						
<i>Trichodesmium</i>			52	15	10 (6.5–16)	71 (19)		
Heterocystous CYN			63	11	0.02 (0.02–0.03)	0.42 (0.12)		
Total					10 (6.6–16)	71 (16)	0.91 (0.58–1.4)	6.3 (1.5)
South Pacific Ocean	40–0° S	72						
<i>Trichodesmium</i>			4	1	0.11 (0.09–0.13)	0.09 (0.03)		
Heterocystous CYN			16	13	0.06 (0.04–0.11)	0.02 (0.01)		
Total					0.17 (0.13–0.25)	0.10 (0.01)	0.013 (0.009–0.018)	0.007 (0.001)
Indian Ocean	40° S–25° N	62	2	0	190 (160–240)	350 (110)	NQ	NQ
Mediterranean Sea	30–45° N	2.5	ND	ND	ND	ND	ND	ND
Global							2.1 (1.4–3.1)	18 (1.8)

**Table 8.** Estimates of diazotrophic biomass for the global oceans using *nifH*-based data. Same method as Table 6 is used. ND: no data. NQ: not qualified to calculate areal sum because of too few samples.

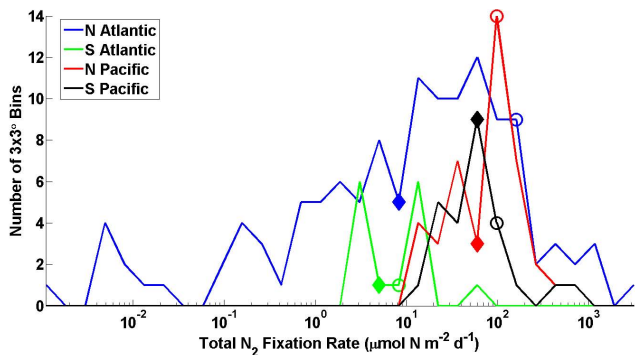
Region	Latitudinal range	Ocean area ( $\times 10^{12}$ m <sup>2</sup> )	<i>n</i> of bins	<i>n</i> of zero-value bins	<i>nifH</i> -based diazotrophic biomass (mg C m <sup>-2</sup> )		Areal Sum (Tg C)	
					Geometric	Arithmetic	Geometric	Arithmetic
North Atlantic Ocean	0–55° N	36	41	6	290 (150–560)	3800 (1200)	11 (5.5–20)	140 (44)
South Atlantic Ocean	40–0° S	27	ND	ND	ND	ND	ND	ND
North Pacific Ocean	0–55° N	89	12	0	170 (78–380)	1700 (630)	15 (6.9–34)	150 (56)
South Pacific Ocean	40–0° S	72	22	0	880 (560–1400)	4200 (1500)	63 (40–100)	310 (110)
Indian Ocean	40° S–25° N	62	4	0	33 (16–69)	81 (63)	NQ	NQ
Mediterranean Sea	30–45° N	2.5	ND	ND	ND	ND	ND	ND
Global							89 (52–150)	590 (70)

than the peak of the distribution. The arithmetic mean in this basin is more than one order of magnitude higher than the geometric mean. In the North and South Pacific, the N<sub>2</sub> fixation rates only differ in 2 orders of magnitude, without any data lower than 10  $\mu\text{mol N m}^{-2} \text{d}^{-1}$  (Fig. 10). Thus, the difference between the geometric and arithmetic means is smaller in the Pacific than in the North Atlantic. Although the geometric mean N<sub>2</sub> fixation rate in the North Atlantic is about one order of magnitude lower than those in the North and South Pacific, the peaks of the distributions are actually much closer in these basins (Fig. 10; also see Table 6). Compared to the North Atlantic, the Pacific is not intensively sampled for N<sub>2</sub> fixation. Thus, we need more samples with wider spatial coverage from the Pacific to better evaluate the global N<sub>2</sub> fixation rate.

Based on depth-integrated data, the global mean estimate of diazotrophic biomass from cell counts is 1–2 orders of magnitude lower than that from *nifH*-based abundances (Tables 7 and 8), which are mostly contributed by the dom-

inant taxon, *Trichodesmium* (Table 5). But considering the log-normal distributions, difference of one or two orders of magnitude is relatively small compared to the span of the distribution especially when the number of samples is limited (see Fig. 7 as an example). More samples are needed in order to confidently validate the usage of *nifH* data for estimating diazotrophic abundances and the assumption of 1 : 1 conversion from *nifH* to cell densities.

The above global estimates are based on default values of the biomass conversion factors. In order to show the effects of the variation of the biomass conversion factors, we have applied the upper and lower bounds of these conversion factors for all diazotrophic subtypes in order to revise global diazotrophic biomass estimates following the same procedure described above. The estimated global geometric mean diazotrophic biomass can vary by about  $\pm 70\%$  from default estimates, ranging in 0.8–3.5 Tg C (default 2.1 Tg C) based on cell count data, or in 28–170 Tg C (default 89 Tg C) based on *nifH*-based data (Table 2).



**Figure 10.** Histograms of depth-integrated total  $N_2$  fixation rates (after binned on  $3^\circ \times 3^\circ$  grid) in North Atlantic (blue), South Atlantic (green), North Pacific (red) and South Pacific (black). For each basin, filled diamond and empty circle mark the location of the geometric mean and the arithmetic mean, respectively.

#### 4 Conclusions and recommendations for use

The first global database of oceanic diazotrophic measurements has been constructed with sub-databases for  $N_2$  fixation rates, cell counts for diazotrophs and *nifH*-based abundances. This database provides useful information on the spatial patterns of  $N_2$  fixation and diazotrophic biomass in the world ocean. Depth-integrated values were used to conduct a first-order estimate of the global  $N_2$  fixation rate and diazotrophic biomass. Spatial coverage is the main limitation of the database, as there are still vast oceanic areas where diazotrophic activity and biomass have never been measured. For example, measurements in the Indian Ocean, the South Atlantic and the subtropical southern Pacific are a high priority in this regard. A finer understanding of the range and applicability of fixed biomass conversion factors is another concern. Although careful analyses have been conducted to derive a reliable global biomass, our estimates may vary by  $\pm 70\%$  depending on the conversion factor. More direct elemental analyses are required to further narrow the range of biomass conversion factors. Relatively higher abundances tend to be found from *nifH*-based data than from the cell count data. The estimated global mean diazotrophic biomass from *nifH*-based abundances is 1–2 orders of magnitude higher than that derived from cell counts. However, in the well-sampled region of the North Atlantic, the *Trichodesmium* abundances estimated from these two types of data match each other better. Further investigations and more data are needed to identify what causes the discrepancies between these two estimates. However, it is clear that microscopic cell counts cannot, at present, account for the UCYN-A, which is very abundant in the ocean.

With the recent findings that most of the previous  $N_2$  fixation rates have been underestimated when the  $^{15}N_2$  tracer was added to the incubations as gas bubbles, it can be expected that higher  $N_2$  fixation rates will be reported in the

near future using the new method by adding  $^{15}N_2$  in a dissolved form. Each data record needs to be well documented for the applied  $^{15}N_2$  method, so that future users can identify which data records are likely underestimated.

The database is stored permanently at PANGAEA and can be easily accessed by users. It will be routinely updated with new measurements. The database can be used to study the level of oceanic  $N_2$  fixation activity and its temporal and spatial variations on local, regional and global scales, as well as for constraining the relative contribution of new nitrogen inputs from  $N_2$  fixation and other sources. The diazotrophic biomass data, along with data from other functional groups, can be used to study phytoplankton community structure. The database can also be used to validate geochemical estimates of  $N_2$  fixation and to parameterize and validate biogeochemical models, keeping in mind that  $N_2$  fixation rate measurements have been underestimating the true  $N_2$  fixation rates.

**Supplementary material related to this article is available online at:** <http://www.earth-syst-sci-data.net/4/47/2012/essd-4-47-2012-supplement.pdf>.

**Acknowledgements.** Y.-W.L., S.C.D., D.B., M.J.C., D.M.K., P.H.M., K.A.T., A.E.W. and J.P.Z. are funded by C-MORE project (NSF EF-0424599). L.A.A. and D.J.M. gratefully acknowledge support of NSF OCE-0925284. M.B. acknowledges funding from projects CAIBEX (CTM2007-66408-CO2-O2) and CYFOD (CTM2008-00915-E), and FPI fellowship (BES-2008-006985) by MICINN (Spain). S.B. is funded by IRD and data are contributions of the BOUM experiment and ISOFeRIX Project of the French national LEFE-CYBER program, the European IP SESAME and the international IMBER project. M.J.C., R.M.L. and D.M.K. acknowledge NSF OCE OCE-08-50827 for supporting  $N_2$  fixation measurements at Station ALOHA. R.J.L. and J.L. are funded by the German BMBF project Sopran. D.M.K., P.H.M., K.A.T. and J.P.Z., are funded by the Gordon and Betty Moore Foundation. C.M.M. is funded by NERC, UK. B.M.-C. was supported by the Ramon y Cajal program from the Spanish Ministry of Science and Technology. A.P.R. received funding through NERC OCEANS2025. A.E.W. acknowledges funding from NSF OCE-0726422 (to F. Pahl, OSU). We thank Nicolas Gruber and another anonymous referee for their constructive reviews.

Edited by: S. Pesant

#### References

- Arin, L., Berdalet, E., Marrasé, C., Estrada, M., Guixa-Boixereu, N., and Dolan, J.: Particulate DNA and protein relative to microorganism biomass and detritus in the Catalano-Balearic Sea (NW Mediterranean) during stratification, *J. Plankton Res.*, 21, 1299–1316, 1999.
- Bailiff, M. D. and Karl, D. M.: Dissolved and Particulate DNA Dynamics during a Spring Bloom in the Antarctic Peninsula Region,

- 1986–87, *Deep-Sea Res.*, 38, 1077–1095, doi:10.1016/0198-0149(91)90097-Y, 1991.
- Bar-Zeev, E., Yogev, T., Man-Aharonovich, D., Kress, N., Herut, B., Beja, O., and Berman-Frank, I.: Seasonal dynamics of the endosymbiotic, nitrogen-fixing cyanobacterium *Richelia intracellularis* in the eastern Mediterranean Sea, *Isme J.*, 2, 911–923, doi:10.1038/ismej.2008.56, 2008.
- Benavides, M., Agawin, N. S. R., Arístegui, J., Ferriol, P., and Stal, L. J.: Nitrogen fixation by *Trichodesmium* and small diazotrophs in the subtropical northeast Atlantic, *Aquat. Microb. Ecol.*, 65, 43–53, doi:10.3354/ame01534, 2011.
- Bonnet, S., Guieu, C., Bruyant, F., Prášil, O., Van Wambeke, F., Raimbault, P., Moutin, T., Grob, C., Gorbunov, M. Y., Zehr, J. P., Masquelier, S. M., Garczarek, L., and Claustre, H.: Nutrient limitation of primary productivity in the Southeast Pacific (BIOSEPE cruise), *Biogeosciences*, 5, 215–225, doi:10.5194/bg-5-215-2008, 2008.
- Bonnet, S., Biegala, I. C., Dutrieux, P., Slemmons, L. O., and Capone, D. G.: Nitrogen fixation in the western equatorial Pacific: Rates, diazotrophic cyanobacterial size class distribution, and biogeochemical significance, *Global Biogeochem. Cy.*, 23, GB3012, doi:10.1029/2008GB003439, 2009.
- Bonnet, S., Grosso, O., and Moutin, T.: Planktonic dinitrogen fixation along a longitudinal gradient across the Mediterranean Sea during the stratified period (BOUM cruise), *Biogeosciences*, 8, 2257–2267, doi:10.5194/bg-8-2257-2011, 2011.
- Borstad, G. A.: Some aspects of the occurrence and biology of *Trichodesmium* (cyanobacteria) in the western tropical Atlantic near Barbados, West Indies, Doctor of Philosophy, Marine Science Centre, McGill University, Montreal, Quebec, Canada, 234 pp., 1978.
- Boström, K. H., Riemann, L., Zweifel, U. L., and Hagström, Å.: *Nodularia* sp *nifH* gene transcripts in the Baltic Sea proper, *J. Plankton Res.*, 29, 391–399, doi:10.1093/plankt/fbm019, 2007.
- Brzezinski, M. A., Villareal, T. A., and Lipschultz, F.: Silica production and the contribution of diatoms to new and primary production in the central North Pacific, *Mar. Ecol.-Prog. Ser.*, 167, 89–104, doi:10.3354/meps167089, 1998.
- Buitenhuis, E. T., Bednarek, N., Doney, S. C., Dunne, J. P., Leblanc, K., Luo, Y.-W., Moriarty, R., O'Brien, T. D., Schiebel, R., Peloquin, J. A., and Vogt, M.: MAREDAT: Towards a world ocean atlas of marine ecosystem data, *Earth Syst. Sci. Data Discuss.*, in preparation, 2012.
- Campbell, J. W.: The lognormal distribution as a model for bio-optical variability in the sea, *J. Geophys. Res.-Oceans*, 100, 13237–13254, doi:10.1029/95JC00458, 1995.
- Capone, D. G.: Determination of nitrogenase activity in aquatic samples using the acetylene reduction procedure, in: *Handbook of Methods in Aquatic Microbial Ecology*, edited by: Kemp, P. F., Cole, J. J., Sherr, B. F., and Sherr, E. B., Lewis Publishers, Boca Raton, Fla, 621–631, 1993.
- Capone, D. G.: The marine nitrogen cycle, in: *Microbial ecology of the ocean*, edited by: Kirchman, D., Wiley-Liss, New York, 455–493, 2000.
- Capone, D. G., Zehr, J. P., Paerl, H. W., Bergman, B., and Carpenter, E. J.: *Trichodesmium*, a globally significant cyanobacterium, *Science*, 276, 1221–1229, doi:10.1126/science.276.5316.1221, 1997.
- Capone, D. G., Subramaniam, A., Montoya, J. P., Voss, M., Humborg, C., Johansen, A. M., Siefert, R. L., and Carpenter, E. J.: An extensive bloom of the N<sub>2</sub>-fixing cyanobacterium *Trichodesmium erythraeum* in the central Arabian Sea, *Mar. Ecol.-Prog. Ser.*, 172, 281–292, doi:10.3354/meps172281, 1998.
- Capone, D. G., Burns, J. A., Montoya, J. P., Subramaniam, A., Mahaffey, C., Gunderson, T., Michaels, A. F., and Carpenter, E. J.: Nitrogen fixation by *Trichodesmium* spp.: An important source of new nitrogen to the tropical and subtropical North Atlantic Ocean, *Global Biogeochem. Cy.*, 19, GB2024, doi:10.1029/2004GB002331, 2005.
- Carpenter, E. J.: Nitrogen fixation by marine *Oscillatoria* (*Trichodesmium*) in the world's oceans, in: *Nitrogen in the Marine Environment*, edited by: Carpenter, E. J. and Capone, D. G., Academic Press, New York, 65–103, 1983.
- Carpenter, E. J. and Capone, D. G.: Nitrogen Fixation in the Marine Environment, in: *Nitrogen in the Marine Environment*, edited by: Capone, D. G., Bronk, D. A., Mulholland, M. R., and Carpenter, E. J., Academic Press, London, 1197–1218, 2008.
- Carpenter, E. J. and Price, C. C.: Nitrogen fixation, distribution, and production of *Oscillatoria* (*Trichodesmium*) Spp. in the western Sargasso and Caribbean Seas, *Limnol. Oceanogr.*, 22, 60–72, 1977.
- Carpenter, E. J. and Romans, K.: Major role of the cyanobacterium *Trichodesmium* in nutrient cycling in the North Atlantic Ocean, *Science*, 254, 1356–1358, doi:10.1126/science.254.5036.1356, 1991.
- Carpenter, E. J., Subramaniam, A., and Capone, D. G.: Biomass and primary productivity of the cyanobacterium, *Trichodesmium* spp., in the tropical N Atlantic Ocean, *Deep-Sea Res. Pt. I*, 51, 173–203, doi:10.1016/j.dsr.2003.10.006, 2004.
- Chang, J., Chiang, K. P., and Gong, G. C.: Seasonal variation and cross-shelf distribution of the nitrogen-fixing cyanobacterium, *Trichodesmium*, in southern East China Sea, *Cont. Shelf Res.*, 20, 479–492, doi:10.1016/S0278-4343(99)00082-5, 2000.
- Chen, Y. L. L., Chen, H. Y., and Lin, Y. H.: Distribution and downward flux of *Trichodesmium* in the South China Sea as influenced by the transport from the Kuroshio Current, *Mar. Ecol.-Prog. Ser.*, 259, 47–57, doi:10.3354/meps259047, 2003.
- Chen, Y. L. L., Chen, H. Y., Tuo, S. H., and Ohki, K.: Seasonal dynamics of new production from *Trichodesmium* N<sub>2</sub> fixation and nitrate uptake in the upstream Kuroshio and South China Sea basin, *Limnol. Oceanogr.*, 53, 1705–1721, doi:10.4319/lo.2008.53.5.1705, 2008.
- Chen, Y. L. L., Tuo, S. H., and Chen, H. Y.: Co-occurrence and transfer of fixed nitrogen from *Trichodesmium* spp. to diatoms in the low-latitude Kuroshio Current in the NW Pacific, *Mar. Ecol.-Prog. Ser.*, 421, 25–38, doi:10.3354/meps08908, 2011.
- Church, M. J., Jenkins, B. D., Karl, D. M., and Zehr, J. P.: Vertical distributions of nitrogen-fixing phylotypes at Stn ALOHA in the oligotrophic North Pacific Ocean, *Aquat. Microb. Ecol.*, 38, 3–14, doi:10.3354/ame038003, 2005a.
- Church, M. J., Short, C. M., Jenkins, B. D., Karl, D. M., and Zehr, J. P.: Temporal patterns of nitrogenase gene (*nifH*) expression in the oligotrophic North Pacific Ocean, *Appl. Environ. Microbiol.*, 71, 5362–5370, doi:10.1128/Aem.71.9.5362-5370.2005, 2005b.
- Church, M. J., Mahaffey, C., Letelier, R. M., Lukas, R., Zehr, J. P., and Karl, D. M.: Physical forcing of nitrogen fixation and diazotroph community structure in the North

- Pacific subtropical gyre, *Global Biogeochem. Cy.*, 23, GB2020, doi:10.1029/2008gb003418, 2009.
- Deutsch, C., Sarmiento, J. L., Sigman, D. M., Gruber, N., and Dunne, J. P.: Spatial coupling of nitrogen inputs and losses in the ocean, *Nature*, 445, 163–167, doi:10.1038/nature05392, 2007.
- Doney, S. C., Lima, I., Moore, J. K., Lindsay, K., Behrenfeld, M. J., Westberry, T. K., Mahowald, N., Glover, D. M., and Takahashi, T.: Skill metrics for confronting global upper ocean ecosystem-biogeochemistry models against field and remote sensing data, *J. Marine Syst.*, 76, 95–112, doi:10.1016/j.jmarsys.2008.05.015, 2009.
- Dong, J. D., Zhang, Y. Y., Wang, Y. S., Zhang, S., and Wang, H. K.: Spatial and seasonal variations of Cyanobacteria and their nitrogen fixation rates in Sanya Bay, South China Sea, *Sci. Mar.*, 72, 239–251, 2008.
- Dore, J. E., Brum, J. R., Tupas, L. M., and Karl, D. M.: Seasonal and interannual variability in sources of nitrogen supporting export in the oligotrophic subtropical North Pacific Ocean, *Limnol. Oceanogr.*, 47, 1595–1607, doi:10.4319/lo.2002.47.6.1595, 2002.
- Dortch, Q., Roberts, T. L., Clayton, J. R., and Ahmed, S. I.: RNA/DNA ratios and DNA concentrations as indicators of growth rate and biomass in planktonic marine organisms, *Mar. Ecol.-Prog. Ser.*, 13, 61–71, 1983.
- Dugdale, R. C., Goering, J. J., and Ryther, J. H.: Nitrogen fixation in the Sargasso Sea, *Deep-Sea Res.*, 7, 298–300, doi:10.1016/0146-6313(61)90051-X, 1961.
- Falcón, L. I., Carpenter, E. J., Cipriano, F., Bergman, B., and Capone, D. G.: N<sub>2</sub> fixation by unicellular bacterioplankton from the Atlantic and Pacific oceans: Phylogeny and in situ rates, *Appl. Environ. Microbiol.*, 70, 765–770, doi:10.1128/Aem.70.2.765-770.2004, 2004.
- Farnelid, H., Andersson, A. F., Bertilsson, S., Abu Al-Soud, W., Hansen, L. H., Sorensen, S., Steward, G. F., Hagstrom, A., and Riemann, L.: Nitrogenase Gene Amplicons from Global Marine Surface Waters Are Dominated by Genes of Non-Cyanobacteria, *Plos One*, 6, e19223, doi:10.1371/journal.pone.0019223, 2011.
- Fernández, A., Mouriño-Carballido, B., Bode, A., Varela, M., and Marañón, E.: Latitudinal distribution of *Trichodesmium* spp. and N<sub>2</sub> fixation in the Atlantic Ocean, *Biogeosciences*, 7, 3167–3176, doi:10.5194/bg-7-3167-2010, 2010.
- Fernández, C., Farías, L., and Ulloa, O.: Nitrogen Fixation in Denitrified Marine Waters, *Plos One*, 6, e20539, doi:10.1371/journal.pone.0020539, 2011.
- Fernández, A., Graña, R., Mouriño-Carballido, B., Bode, A., Varela, M., Domínguez, J. F., Escánez, J., de Armas, D., and Marañón, E.: Community N<sub>2</sub> fixation and *Trichodesmium* spp. abundance along longitudinal gradients in the Eastern Subtropical North Atlantic, in preparation, 2012.
- Fong, A. A., Karl, D. M., Lukas, R., Letelier, R. M., Zehr, J. P., and Church, M. J.: Nitrogen fixation in an anticyclonic eddy in the oligotrophic North Pacific Ocean, *ISME J.*, 2, 663–676, doi:10.1038/ismej.2008.22, 2008.
- Foster, R. A. and O'Mullan, G. D.: Nitrogen-fixing and nitrifying Symbioses in the marine environment, in: *Nitrogen in the Marine Environment*, edited by: Capone, D. G., Bronk, D. A., Mulholland, M. R., and Carpenter, E. J., Academic Press, London, 1197–1218, 2008.
- Foster, R. A. and Zehr, J. P.: Characterization of diatom-cyanobacteria symbioses on the basis of *nifH*, *hetR* and 16S rRNA sequences, *Environ. Microb.*, 8, 1913–1925, doi:10.1111/j.1462-2920.2006.01068.x, 2006.
- Foster, R. A., Subramaniam, A., Mahaffey, C., Carpenter, E. J., Capone, D. G., and Zehr, J. P.: Influence of the Amazon River plume on distributions of free-living and symbiotic cyanobacteria in the western tropical north Atlantic Ocean, *Limnol. Oceanogr.*, 52, 517–532, doi:10.4319/lo.2007.52.2.0517, 2007.
- Foster, R. A., Paytan, A., and Zehr, J. P.: Seasonality of N<sub>2</sub> fixation and *nifH* gene diversity in the Gulf of Aqaba (Red Sea), *Limnol. Oceanogr.*, 54, 219–233, doi:10.4319/lo.2009.54.1.0219, 2009a.
- Foster, R. A., Subramaniam, A., and Zehr, J. P.: Distribution and activity of diazotrophs in the Eastern Equatorial Atlantic, *Environ. Microbiol.*, 11, 741–750, doi:10.1111/j.1462-2920.2008.01796.x, 2009b.
- 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, *ISME J.*, 5, 1484–1493, doi:10.1038/ismej.2011.26, 2011.
- Galloway, J. N., Dentener, F. J., Capone, D. G., Boyer, E. W., Howarth, R. W., Seitzinger, S. P., Asner, G. P., Cleveland, C. C., Green, P. A., Holland, E. A., Karl, D. M., Michaels, A. F., Porter, J. H., Townsend, A. R., and Vorosmarty, C. J.: Nitrogen cycles: past, present, and future, *Biogeochemistry*, 70, 153–226, doi:10.1007/s10533-004-0370-0, 2004.
- Gandhi, N., Singh, A., Prakash, S., Ramesh, R., Raman, M., Sheshshayee, M. S., and Shetye, S.: First direct measurements of N<sub>2</sub> fixation during a *Trichodesmium* bloom in the eastern Arabian Sea, *Global Biogeochem. Cy.*, 25, Gb4014, doi:10.1029/2010gb003970, 2011.
- Garcia, N., Raimbault, P., and Sandroni, V.: Seasonal nitrogen fixation and primary production in the Southwest Pacific: nanoplankton diazotrophy and transfer of nitrogen to picoplankton organisms, *Mar. Ecol.-Prog. Ser.*, 343, 25–33, doi:10.3354/Meps06882, 2007.
- Glover, D. M., Jenkins, W. J., and Doney, S. C.: *Modeling Methods for Marine Science*, Cambridge University Press, Cambridge, UK, 2011.
- Goebel, N. L., Edwards, C. A., Carter, B. J., Achilles, K. M., and Zehr, J. P.: Growth and carbon content of three different-sized diazotrophic cyanobacteria observed in the subtropical North Pacific, *J. Phycol.*, 44, 1212–1220, doi:10.1111/j.1529-8817.2008.00581.x, 2008.
- Goebel, N. L., Turk, K. A., Achilles, K. M., Paerl, R., Hewson, I., Morrison, A. E., Montoya, J. P., Edwards, C. A., and Zehr, J. P.: Abundance and distribution of major groups of diazotrophic cyanobacteria and their potential contribution to N<sub>2</sub> fixation in the tropical Atlantic Ocean, *Environ. Microbiol.*, 12, 3272–3289, doi:10.1111/j.1462-2920.2010.02303.x, 2010.
- Goering, J. J., Dugdale, R. C., and Menzel, D. W.: Estimates of *in situ* rates of nitrogen uptake by *Trichodesmium* sp. in the tropical Atlantic Ocean, *Limnol. Oceanogr.*, 11, 614–620, 1966.
- Gómez, F., Furuya, K., and Takeda, S.: Distribution of the cyanobacterium *Richelia intracellularis* as an epiphyte of the diatom *Chaetoceros compressus* in the western Pacific Ocean, *J. Plankton Res.*, 27, 323–330, doi:10.1093/plankt/fbi007, 2005.
- Grabowski, M. N. W., Church, M. J., and Karl, D. M.: Nitrogen fixation rates and controls at Stn ALOHA, *Aquat. Microb. Ecol.*,



- 52, 175–183, doi:10.3354/Ame01209, 2008.
- Großkopf, T., Mohr, W., Baustian, T., Schunck, H., Gill, D., Kuypers, M. M. M., Lavik, G., Schmitz, R. A., R. Wallace, D. W., and LaRoche, J.: Doubling of marine dinitrogen-fixation rates based on direct measurements, *Nature*, 488, 361–364, doi:10.1038/nature11338, 2012.
- Gruber, N.: The Marine Nitrogen Cycle: Overview and Challenges, in: *Nitrogen in the marine environment*, 2nd Edn., edited by: Capone, D. G., Bronk, D. A., Mulholland, M. R., and Carpenter, E. J., Elsevier, Amsterdam, 1–50, 2008.
- Gruber, N. and Sarmiento, J. L.: Global patterns of marine nitrogen fixation and denitrification, *Global Biogeochem. Cy.*, 11, 235–266, 1997.
- Gruber, N. and Sarmiento, J. L.: Large-scale biogeochemical/physical interactions in elemental cycles, in: *The Sea. Biological-Physical Interactions in the Sea*, edited by: Robinson, A. R., McCarthy, J. J., and Rothschild, B. J., John Wiley & Sons, New York City, 2002.
- Gundersen, K. R., Corbin, J. S., Hanson, C. L., Hanson, M. L., Hanson, R. B., Russell, D. J., Stollar, A., and Yamadas, O.: Structure and biological dynamics of the oligotrophic ocean photic zone off the Hawaiian Islands, *Pac. Sci.*, 30, 45–68, 1976.
- Hegde, S., Anil, A. C., Patil, J. S., Mitbavkar, S., Krishnamurthy, V., and Gopalakrishna, V. V.: Influence of environmental settings on the prevalence of *Trichodesmium* spp. in the Bay of Bengal, *Mar. Ecol.-Prog. Ser.*, 356, 93–101, doi:10.3354/Meps07259, 2008.
- Holm-Hansen, O.: Determination of microbial biomass in ocean profiles, *Limnol. Oceanogr.*, 14, 740–747, 1969.
- Holm-Hansen, O., Sutcliffe Jr., W. H., and Sharp, J.: Measurement of Deoxyribonucleic Acid in the Ocean and its Ecological Significance, *Limnol. Oceanogr.*, 13, 507–514, 1968.
- Hulburt, E. M.: Phytoplankton in the Southwestern Sargasso Sea and North Equatorial current, *Limnol. Oceanogr.*, 7, 307–315, 1962.
- Hulburt, E. M.: The distribution of phytoplankton, and its relationship to hydrography, between southern New England and Venezuela, *J. Mar. Res.*, 24, 67–81, 1966.
- Hulburt, E. M.: Phytoplankton observations in the western Caribbean Sea, *B. Mar. Sci.*, 18, 388–399, 1968.
- Hynes, A. M., Webb, E. A., Doney, S. C., and Waterbury, J. B.: Comparison of cultured *Trichodesmium* (Cyanophyceae) with species characterized from the field, *J. Phycol.*, 48, 196–210, doi:10.1111/j.1529-8817.2011.01096.x, 2012.
- Janson, S., Wouters, J., Bergman, B., and Carpenter, E. J.: Host specificity in the *Richelia*-diatom symbiosis revealed by hetR gene sequence analysis, *Environ. Microbiol.*, 1, 431–438, doi:10.1038/ismej.2011.26, 1999.
- Karl, D., Michaels, A., Bergman, B., Capone, D., Carpenter, E., Letelier, R., Lipschitz, F., Paerl, H., Sigman, D., and Stal, L.: Dinitrogen fixation in the world's oceans, *Biogeochemistry*, 57/58, 47–98, doi:10.1023/A:1015798105851, 2002.
- Karl, D. M., Letelier, R., Tupas, L., Dore, J., Christian, J., and Hebel, D.: The role of nitrogen fixation in biogeochemical cycling in the subtropical North Pacific Ocean, *Nature*, 388, 533–538, 1997.
- Kitajima, S., Furuya, K., Hashihama, F., Takeda, S., and Kanda, J.: Latitudinal distribution of diazotrophs and their nitrogen fixation in the tropical and subtropical western North Pacific, *Limnol. Oceanogr.*, 54, 537–547, doi:10.4319/lo.2009.54.2.0537, 2009.
- Koch, A. L.: Logarithm in Biology. 1. Mechanisms Generating Log-Normal Distribution Exactly, *J. Theor. Biol.*, 12, 276–290, doi:10.1016/0022-5193(66)90119-6, 1966.
- Kong, L., Jing, H., Kataoka, T., Sun, J., and Liu, H.: Phylogenetic diversity and spatio-temporal distribution of nitrogenase genes (*nifH*) in the northern South China Sea, *Aquat. Microb. Ecol.*, 65, 15–27, doi:10.3354/ame01531, 2011.
- Kromkamp, J., De Bie, M., Goosen, N., Peene, J., Van Rijswijk, P., Sinke, J., and Duineveld, G. C. A.: Primary production by phytoplankton along the Kenyan coast during the SE monsoon and November intermonsoon 1992, and the occurrence of *Trichodesmium*, *Deep-Sea Res. Pt. II*, 44, 1195–1212, doi:10.1016/S0967-0645(97)00015-5, 1997.
- Langlois, R. J., LaRoche, J., and Raab, P. A.: Diazotrophic diversity and distribution in the tropical and subtropical Atlantic ocean, *Appl. Environ. Microbiol.*, 71, 7910–7919, doi:10.1128/Aem.71.12.7910-7919.2005, 2005.
- Langlois, R. J., Hümmel, D., and LaRoche, J.: Abundances and distributions of the dominant *nifH* phylotypes in the Northern Atlantic Ocean, *Appl. Environ. Microbiol.*, 74, 1922–1931, doi:10.1128/Aem.01720-07, 2008.
- LaRoche, J. and Breitbart, E.: Importance of the diazotrophs as a source of new nitrogen in the ocean, *J. Sea Res.*, 53, 67–91, doi:10.1016/j.seares.2004.05.005, 2005.
- Letelier, R. M. and Karl, D. M.: Role of *Trichodesmium* spp. in the productivity of the subtropical North Pacific Ocean, *Mar. Ecol.-Prog. Ser.*, 133, 263–273, doi:10.3354/meps133263, 1996.
- Limpert, E., Stahel, W. A., and Abbt, M.: Log-normal distributions across the sciences: Keys and clues, *Bioscience*, 51, 341–352, doi:10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2, 2001.
- Lugomela, C., Lyimo, T. J., Bryceson, I., Semesi, A. K., and Bergman, B.: *Trichodesmium* in coastal waters of Tanzania: diversity, seasonality, nitrogen and carbon fixation, *Hydrobiologia*, 477, 1–13, doi:10.1023/A:1021017125376, 2002.
- Mague, T. H., Weare, N. M., and Holm-Hansen, O.: Nitrogen fixation in the North Pacific Ocean, *Mar. Biol.*, 24, 109–119, 1974.
- Mague, T. H., Mague, F. C., and Holm-Hansen, O.: Physiology and chemical composition of nitrogen-fixing phytoplankton in the central North Pacific Ocean, *Mar. Biol.*, 41, 213–227, 1977.
- Marumo, R. and Asaoka, O.: Distribution of pelagic blue-green algae in the North Pacific Ocean, *Journal of the Oceanographical Society of Japan*, 30, 77–85, doi:10.1007/bf02112896, 1974.
- Mazard, S. L., Fuller, N. J., Orcutt, K. M., Bridle, O., and Scanlan, D. J.: PCR analysis of the distribution of unicellular cyanobacterial diazotrophs in the Arabian Sea, *Appl. Environ. Microbiol.*, 70, 7355–7364, doi:10.1128/Aem.70.12.7355-7364.2004, 2004.
- McCarthy, J. J. and Carpenter, E. J.: *Oscillatoria* (*Trichodesmium*) *thiebautii* (cyanophyta) in the central North Atlantic Ocean, *J. Phycol.*, 15, 75–82, doi:10.1111/j.1529-8817.1979.tb02965.x, 1979.
- Mohr, W., Grosskopf, T., Wallace, D. W. R., and LaRoche, J.: Methodological underestimation of oceanic nitrogen fixation rates, *Plos One*, 5, e12583, doi:10.1371/journal.pone.0012583, 2010.
- Moisander, P. H., Beinart, R. A., Voss, M., and Zehr, J. P.: Diversity and abundance of diazotrophic microorganisms in the South China Sea during intermonsoon, *ISME J.*, 2, 954–967, doi:10.1038/ismej.2008.51, 2008.

- Moisander, P. H., Beinart, R. A., Hewson, I., White, A. E., Johnson, K. S., Carlson, C. A., Montoya, J. P., and Zehr, J. P.: Unicellular cyanobacterial distributions broaden the oceanic N<sub>2</sub> fixation domain, *Science*, 327, 1512–1514, doi:10.1126/science.1185468, 2010.
- Montoya, J. P., Voss, M., Kahler, P., and Capone, D. G.: A simple, high-precision, high-sensitivity tracer assay for N<sub>2</sub> fixation, *Appl. Environ. Microbiol.*, 62, 986–993, 1996.
- Montoya, J. P., Houll, C. M., Zehr, J. P., Hansen, A., Villareal, T. A., and Capone, D. G.: High rates of N<sub>2</sub> fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean, *Nature*, 430, 1027–1032, doi:10.1038/nature02824, 2004.
- Moore, C. M., Mills, M. M., Achterberg, E. P., Geider, R. J., LaRoche, J., Lucas, M. I., McDonagh, E. L., Pan, X., Poulton, A. J., Rijkenberg, M. J. A., Suggett, D. J., Ussher, S. J., and Woodward, E. M. S.: Large-scale distribution of Atlantic nitrogen fixation controlled by iron availability, *Nat. Geosci.*, 2, 867–871, doi:10.1038/ngeo667, 2009.
- Mouriño-Carballido, B., Graña, R., Fernández, A., Bode, A., Varela, M., Domínguez, J. F., Escánez, J., Armas, D. D., and Marañón, E.: Importance of N<sub>2</sub> fixation vs. nitrate eddy diffusion along a latitudinal transect in the Atlantic Ocean, *Limnol. Oceanogr.*, 56, 999–1007, doi:10.4319/lo.2011.56.3.0999, 2011.
- Mulholland, M. R.: The fate of nitrogen fixed by diazotrophs in the ocean, *Biogeosciences*, 4, 37–51, doi:10.5194/bg-4-37-2007, 2007.
- Mulholland, M. R., Bronk, D. A., and Capone, D. G.: Dinitrogen fixation and release of ammonium and dissolved organic nitrogen by *Trichodesmium* IMS101, *Aquat. Microb. Ecol.*, 37, 85–94, doi:10.3354/ame037085, 2004.
- Mulholland, M. R., Bernhardt, P. W., Blanco-Garcia, J. L., Mannino, A., Hyde, K., Mondragon, E., Turk, K., Moisander, P. H., and Zehr, J. P.: Rates of dinitrogen fixation and the abundance of diazotrophs in North American coastal waters between Cape Hatteras and Georges Bank, *Limnol. Oceanogr.*, 57, 1067–1083, doi:10.4319/lo.2012.57.4.1067, 2012.
- Needoba, J. A., Foster, R. A., Sakamoto, C., Zehr, J. P., and Johnson, K. S.: Nitrogen fixation by unicellular diazotrophic cyanobacteria in the temperate oligotrophic North Pacific Ocean, *Limnol. Oceanogr.*, 52, 1317–1327, doi:10.4319/lo.2007.52.4.1317, 2007.
- Orcutt, K. M., Lipschultz, F., Gundersen, K., Arimoto, R., Michaels, A. F., Knap, A. H., and Gallon, J. R.: A seasonal study of the significance of N<sub>2</sub> fixation by *Trichodesmium* spp. at the Bermuda Atlantic Time-series Study (BATS) site, *Deep-Sea Res. Pt. II*, 48, 1583–1608, doi:10.1016/S0967-0645(00)00157-0, 2001.
- Postgate, J. R.: *The Fundamentals of Nitrogen Fixation*, Cambridge University Press, New York, 1998.
- Poulton, A. J., Stinchcombe, M. C., and Quartly, G. D.: High numbers of *Trichodesmium* and diazotrophic diatoms in the southwest Indian Ocean, *Geophys. Res. Lett.*, 36, L15610, doi:10.1029/2009GL039719, 2009.
- Rahav, E., Herut, B., Stambler, N., Bar-Zeev, E., Gertman, I., Mulholland, R. M., and Berman-Frank, I.: Summertime dinitrogen fixation in a gyre and eddy of the eastern Mediterranean Sea, *Geophys. Res. Lett.*, in review, 2011.
- Rahav, E., Herut, B., Levi, A., Mulholland, M., and Berman-Frank, I.: Controls on surface dinitrogen fixation across the Mediterranean during springtime, in preparation, 2012a.
- Rahav, E., Herut, B., Levi, A., Mulholland, M., and Berman-Frank, I.: Dinitrogen fixation at the stratified and mixed periods of the Gulf of Aqaba (Red Sea), in preparation, 2012b.
- Raimbault, P. and Garcia, N.: Evidence for efficient regenerated production and dinitrogen fixation in nitrogen-deficient waters of the South Pacific Ocean: impact on new and export production estimates, *Biogeosciences*, 5, 323–338, doi:10.5194/bg-5-323-2008, 2008.
- Reddy, K. J., Haskell, J. B., Sherman, D. M., and Sherman, L. A.: Unicellular, Aerobic Nitrogen-Fixing Cyanobacteria of the Genus *Cyanothece*, *J. Bacteriol.*, 175, 1284–1292, 1993.
- Rees, A. P., Law, C. S., and Woodward, E. M. S.: High rates of nitrogen fixation during an in-situ phosphate release experiment in the Eastern Mediterranean Sea, *Geophys. Res. Lett.*, 33, L10607, doi:10.1029/2006GL025791, 2006.
- Rees, A. P., Gilbert, J. A., and Kelly-Gerreyn, B. A.: Nitrogen fixation in the western English Channel (NE Atlantic Ocean), *Mar. Ecol.-Prog. Ser.*, 374, 7–12, doi:10.3354/meps07771, 2009.
- Riemann, L., Farnelid, H., and Steward, G. F.: Nitrogenase genes in non-cyanobacterial plankton: prevalence, diversity and regulation in marine waters, *Aquat. Microb. Ecol.*, 61, 225–237, doi:10.3354/ame01431, 2010.
- Sandroni, V., Raimbault, P., Migon, C., Garcia, N., and Gouze, E.: Dry atmospheric deposition and diazotrophy as sources of new nitrogen to northwestern Mediterranean oligotrophic surface waters, *Deep-Sea Res. Pt. I*, 54, 1859–1870, doi:10.1016/j.dsr.2007.08.004, 2007.
- Sato, S., Maryse, N. R., and Eskinazi, E.: On the mechanism of red tide of *Trichodesmium* in Recife, northwest Brazil, with some considerations of the relation to the human disease “Tamadare fever”, *Trab. Inst. Oceanogr. Univ. Recife*, 5, 7–49, 1966.
- Scranton, M. I.: Hydrogen cycling in the waters near Bermuda: the role of the nitrogen fixer, *Oscillatoria thiebautii*, *Deep-Sea Res.*, 31, 133–143, 1984.
- Shiozaki, T., Furuya, K., Kodama, T., Kitajima, S., Takeda, S., Takemura, T., and Kanda, J.: New estimation of N<sub>2</sub> fixation in the western and central Pacific Ocean and its marginal seas, *Global Biogeochem. Cy.*, 24, GB1015, doi:10.1029/2009GB003620, 2010.
- Sohm, J. A., Subramaniam, A., Gunderson, T. E., Carpenter, E. J., and Capone, D. G.: Nitrogen fixation by *Trichodesmium* spp. and unicellular diazotrophs in the North Pacific Subtropical Gyre, *J. Geophys. Res.-Biogeo.*, 116, G03002, doi:10.1029/2010JG001513, 2011a.
- Sohm, J. A., Webb, E. A., and Capone, D. G.: Emerging patterns of marine nitrogen fixation, *Nat. Rev. Microbiol.*, 9, 499–508, doi:10.1038/nrmicro2594, 2011b.
- Steven, D. M. and Glombitza, R.: Oscillatory variation of a phytoplankton population in a tropical ocean, *Nature*, 237, 105–107, doi:10.1038/237105a0, 1972.
- Sundström, B. G.: Observations on *Rhizosolenia clevei* Ostensfeld (Bacillariophyceae) and *Richelia intracellularis* Schmidt (Cyanophyceae), *Bot. Mar.*, 27, 345–355, 1984.
- Taniuchi, Y., Chen, Y.-I. L., Chen, H.-Y., Tsai, M.-L., and Ohki, K.: Isolation and characterization of the unicellular diazotrophic cyanobacterium Group C TW3 from the tropical western Pacific Ocean, *Environ. Microbiol.*, 14, 641–654, doi:10.1111/j.1462-2920.2011.02606.x, 2012.

- Tuit, C., Waterbury, J., and Ravizzaz, G.: Diel variation of molybdenum and iron in marine diazotrophic cyanobacteria, *Limnol. Oceanogr.*, 49, 978–990, doi:10.4319/lo.2004.49.4.0978, 2004.
- Turk, K. A., Rees, A. P., Zehr, J. P., Pereira, N., Swift, P., Shelley, R., Lohan, M., Woodward, E. M. S., and Gilbert, J.: Nitrogen fixation and nitrogenase (*nifH*) expression in tropical waters of the eastern North Atlantic, *ISME J.*, 5, 1201–1212, doi:10.1038/ismej.2010.205, 2011.
- Tyrrell, T.: The relative influences of nitrogen and phosphorus on oceanic primary production, *Nature*, 400, 525–531, doi:10.1038/22941, 1999.
- Tyrrell, T., Marañón, E., Poulton, A. J., Bowie, A. R., Harbour, D. S., and Woodward, E. M. S.: Large-scale latitudinal distribution of *Trichodesmium* spp. in the Atlantic Ocean, *J. Plankton Res.*, 25, 405–416, doi:10.1093/plankt/25.4.405, 2003.
- Venrick, E. L.: Distribution and Significance of *Richelia Intracellularis* Schmidt in North Pacific Central Gyre, *Limnol. Oceanogr.*, 19, 437–445, 1974.
- Verity, P. G., Robertson, C. Y., Tronzo, C. R., Andrews, M. G., Nelson, J. R., and Sieracki, M. E.: Relationships between cell volume and the carbon and nitrogen content of marine photosynthetic nanoplankton, *Limnol. Oceanogr.*, 37, 1434–1446, 1992.
- Villareal, T. A.: Division Cycles in the Nitrogen-Fixing *Rhizosolenia* (Bacillariophyceae)-*Richelia* (Nostocaceae) Symbiosis, *Brit. Phycol. J.*, 24, 357–365, doi:10.1080/00071618900650371, 1989.
- Villareal, T. A.: Laboratory Culture and Preliminary Characterization of the Nitrogen-Fixing *Rhizosolenia-Richelia* Symbiosis, *Mar. Ecol.*, 11, 117–132, doi:10.1111/j.1439-0485.1990.tb00233.x, 1990.
- Villareal, T. A.: Marine nitrogen-fixing diatom-cyanobacteria symbioses, in: *Marine Pelagic Cyanobacteria: Trichodesmium and other diazotrophs*, edited by: Carpenter, E. J., Capone, D. G., and Rueter, J. G., NATO ASI Series, 362, Kluwer Academic Publishers, Boston, 163–175, 1992.
- Villareal, T. A., Adornato, L., Wilson, C., and Schoenbaechler, C. A.: Summer blooms of diatom-diazotroph assemblages and surface chlorophyll in the North Pacific gyre: A disconnect, *J. Geophys. Res.-Oceans*, 116, C03001, doi:10.1029/2010jc006268, 2011.
- Watkins-Brandt, K. S., Letelier, R. M., Spitz, Y. H., Church, M. J., Böttjer, D., and White, A. E.: Addition of inorganic or organic phosphorus enhances nitrogen and carbon fixation in the oligotrophic North Pacific, *Mar. Ecol.-Prog. Ser.*, 432, 17–29, doi:10.3354/meps09147, 2011.
- Webb, E. A., Ehrenreich, I. M., Brown, S. L., Valois, F. W., and Waterbury, J. B.: Phenotypic and genotypic characterization of multiple strains of the diazotrophic cyanobacterium, *Crocospaera watsonii*, isolated from the open ocean, *Environ. Microbiol.*, 11, 338–348, doi:10.1111/j.1462-2920.2008.01771.x, 2009.
- White, A. E., Foster, R. A., Benitez-Nelson, C. R., Masqué, P., Verdery, E., Popp, B. N., Arthur, K. E., and Prahl, F. G.: Nitrogen fixation in the Gulf of California and the Eastern Tropical North Pacific, *Prog. Oceanogr.*, submitted, 2012.
- Wilson, S. T., Böttjer, D., Church, M. J., and Karl, D. M.: Comparative assessment of nitrogen fixation methodologies conducted in the oligotrophic North Pacific Ocean, *Appl. Environ. Microbiol.*, 78, 6516–6523, doi:10.1128/AEM.01146-12, 2012.
- Winn, C. D. and Karl, D. M.: Diel Nucleic-Acid Synthesis and Particulate DNA Concentrations – Conflicts with Division Rate Estimates by DNA Accumulation, *Limnol. Oceanogr.*, 31, 637–645, 1986.
- Wu, J. F., Chung, S. W., Wen, L. S., Liu, K. K., Chen, Y. L. L., Chen, H. Y., and Karl, D. M.: Dissolved inorganic phosphorus, dissolved iron, and *Trichodesmium* in the oligotrophic South China Sea, *Global Biogeochem. Cy.*, 17, 1008, doi:10.1029/2002gb001924, 2003.
- Yogev, T., Rahav, E., Bar-Zeev, E., Man-Aharonovich, D., Stambler, N., Kress, N., Beja, O., Mulholland, M. R., Herut, B., and Berman-Frank, I.: Is dinitrogen fixation significant in the Levantine Basin, East Mediterranean Sea?, *Environ. Microbiol.*, 13, 854–871, doi:10.1111/j.1462-2920.2010.02402.x, 2011.
- Zehr, J. P.: Nitrogen fixation by marine cyanobacteria, *Trends Microbiol.*, 19, 162–173, doi:10.1016/j.tim.2010.12.004, 2011.
- Zehr, J. P., Mellon, M. T., and Zani, S.: New nitrogen-fixing microorganisms detected in oligotrophic oceans by amplification of nitrogenase (*nifH*) genes, *Appl. Environ. Microbiol.*, 64, 3444–3450, 1998.
- Zehr, J. P., Waterbury, J. B., Turner, P. J., Montoya, J. P., Omereg, E., Steward, G. F., Hansen, A., and Karl, D. M.: Unicellular cyanobacteria fix N<sub>2</sub> in the subtropical North Pacific Ocean, *Nature*, 412, 635–638, doi:10.1038/35088063, 2001.
- Zehr, J. P., Montoya, J. P., Jenkins, B. D., Hewson, I., Mondragon, E., Short, C. M., Church, M. J., Hansen, A., and Karl, D. M.: Experiments linking nitrogenase gene expression to nitrogen fixation in the North Pacific subtropical gyre, *Limnol. Oceanogr.*, 52, 169–183, doi:10.4319/lo.2007.52.1.0169, 2007.
- Zehr, J. P., Bench, S. R., Carter, B. J., Hewson, I., Niaz, F., Shi, T., Tripp, H. J., and Affourtit, J. P.: Globally Distributed Uncultivated Oceanic N<sub>2</sub>-Fixing Cyanobacteria Lack Oxygenic Photosystem II, *Science*, 322, 1110–1112, doi:10.1126/science.1165340, 2008.