Earth Syst. Sci. Data Discuss., 5, 47–106, 2012
www.earth-syst-sci-data-discuss.net/5/47/2012/
doi:10.5194/essdd-5-47-2012

Open Access Earth System Science Data Discussions

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

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

[1]Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA
[2]Department of Applied Ocean Science and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA
[3]Instituto de Oceanografía y Cambio Global, Universidad de Las Palmas de Gran Canaria, 35017, Las Palmas de Gran Canaria, Spain

[4]Inst. Español de Oceanografía, Centro Oceanográfico de A Coruña, 15080 A Coruña, Spain
[5]IRD-INSU-CNRS, Laboratoire d'Océanographie Physique et Biogéochimique, UMR6535, Centre d'Océanologie de Marseille, Aix Marseille Université, France
[6]Department of Natural Sciences, Linnaeus University, 39182 Kalmar, Sweden
[7]School of Ocean and Earth Science and Technology, University of Hawaii, Honolulu, Hawaii 96822, USA
[8]Department of Biological Sciences and Wrigley Institute for Environmental Studies, University of Southern California, Los Angeles, California 90089, USA
[9]Department of Biological Sciences and Wrigley Institute for Environmental Studies, University of Southern California, Los Angeles, California 90089, USA
[10]Department of Marine Biotechnology and Resources, National Sun Yat-sen University, Kaohsiung 80424, Taiwan
[11]Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, MT 59717, USA
[12]Laboratorio de Ecología Bacteriana, Instituto de Ecología, Universidad Nacional Autónoma de México, Mexic
[13]Departamento de Ecoloxía e Bioloxía Animal, Universidade de Vigo, 36310 Vigo, Spain
[14]Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, 28359 Bremen, Germany
[15]Dept. of Aquatic Bioscience, University of Tokyo, Yayoi, Bunkyo, Tokyo, 113-8657, Japan
[16]Instituto Cavanilles de Biodiversidad y Biología Evolutiva, Universidad de Valencia, P.O. Box 22085, 46071 Valencia, Spain
[17]Department of Marine Science, University of Southern Mississippi, Stennis Space Center, Mississippi 39529, USA
[18]MIT-WHOI Joint Program in Oceanography/Applied Ocean Science and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA
[19]Leibniz Institute for Marine Sciences, Duesternbrooker Weg 20, 24105 Kiel, Germany
[20]College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331, USA
[21]Ocean Sciences Department, University of California, Santa Cruz, California 95064, USA

ESSDD
5, 47–106, 2012

[22]University of Southampton, National Oceanography Centre, Southampton, European Way, Southampton SO14 3ZH, UK
[23]Department of Ocean, Earth & Atmospheric Sciences, Old Dominion University, Norfolk, VA 23529, USA
[24]Division of Environmental and Biomolecular Systems, Oregon Health & Science University, Portland, Oregon 97239, USA
[25]National Oceanography Centre, Waterfront Campus, European Way, Southampton, SO14 3ZH, UK
[26]Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth PL1 3DH, UK
[27]Marine Biological Section, University of Copenhagen, 3000 Helsingør, Denmark
[28]Lamont Doherty Earth Observatory, Palisades, New York 10964, USA
[29]Marine Science Institute, University of Texas at Austin, Port Aransas, Texas 78382, USA
[30]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

Received: 27 December 2011 – Accepted: 5 February 2012 – Published: 13 February 2012

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

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Marine N_2 fixing microorganisms, termed diazotrophs, are a key functional group in marine pelagic ecosystems. The biological fixation of dinitrogen (N_2) 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_2 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. Lower mean N_2 fixation rate was found in the North Atlantic Ocean than the Pacific Ocean. Reporting the geometric mean and the range of one geometric standard error below and above the geometric mean, the pelagic N_2 fixation rate in the global ocean is estimated to be 62 (53–73) Tg N yr^{-1} and the pelagic diazotrophic biomass in the global ocean is estimated to be 4.7 (2.3–9.6) Tg C from cell counts and to 89 (40–200) Tg C from *nifH*-based abundances. Uncertainties related to biomass conversion factors can change the estimate of geometric mean pelagic diazotrophic biomass in the global ocean by about ±70 %. This evolving database can be used to study spatial and temporal distributions and variations of marine N_2 fixation, to validate geochemical estimates and to parameterize and validate biogeochemical models. The database is stored in PANGAEA (http://doi.pangaea.de/10.1594/PANGAEA.774851).

ESSDD
5, 47–106, 2012

1 Introduction

N_2 fixation is the biological conversion of dinitrogen (N_2) gas into two molecules of ammonia by diazotrophic organisms. Over geological time scales, N_2 fixation is important for regulating fixed N concentrations in the ocean and thereby sustain ocean fertility (Tyrrell, 1999). The rate of pelagic N_2 fixation in the contemporary ocean has been estimated to be 100–200 Tg nitrogen (N) yr^{-1}, 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_2-fixing microorganisms in the ocean (Karl et al., 2002; Zehr, 2011). However, non-cyanobacterial prokaryotic plankton may also conduct N_2 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; Fernandez et al., 2011), albeit their relative contribution to global N_2 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; Venrick, 1974; Mague et al., 1977; Villareal, 1990; Foster and O’Mullan, 2008).

ESSDD
5, 47–106, 2012

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. and subsequently, Montoya et al. (2004) measured high rates of N_2 fixation by UCYN in the Pacific Ocean. Three distinct phylogenetic groups have been identified from UCYN, including *Crocosphaera 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., 2011).

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_2 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_2 fixation rate of diazotrophs in the global ocean. This effort is part of the Marine Ecosystem Model Inter-comparison 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

ESSDD
5, 47–106, 2012

are presented in other papers of this special volume. The MAREDAT databases will be used for the future model inter-comparison 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 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_2 fixation rates and *nifH*-based abundances, (3) general characteristics of the datasets including mean N_2 fixation rates and carbon biomass as a function of geographical location and depth, (4) estimates for global N_2 fixation rates and C biomass, and (5) the sensitivity of biomass estimate on the conversion factors. 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 (http://doi.pangaea.de/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_2 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 11 928 data points including three sub-databases: (1) cell counts of diazotrophs with 5191 data points (Table 1a), (2) N_2 fixation rates with 3536 data points (Table 1b) and (3) diazotrophic abundances estimated from *nifH* copy abundances (referred to as *nifH*-based

ESSDD
5, 47–106, 2012

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_2 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_2 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 they are derived values. An additional 408 data points of diazotroph cell counts and N_2 fixation rates that were originally reported as depth-integrated values were also included. These 408 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_2 fixation rates are also provided: when whole seawater N_2 fixation rates are available, they are used as total N_2 fixation rates, otherwise the total N_2 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.

ESSDD
5, 47–106, 2012

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 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., 2011).

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^{-1} and 100 cells trichome^{-1} (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^{-1} 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., unpublished Gulf of Mexico data; Villareal et al., 2011) in which abundances of host diatom *Hemiaulus* and *Rhizosolenia*

ESSDD
5, 47–106, 2012

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 were 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; Villareal, 1992; Janson et al., 1999).

2.1.3 N_2 fixation rates

N_2 fixation rates were measured directly by ^{15}N-labelled N_2 gas ($^{15}N_2$) assimilation (Montoya et al., 1996) or indirectly by the acetylene (C_2H_2) reduction assay (Capone, 1993) (Table 1b). The $^{15}N_2$ assimilation method tracks the conversion of $^{15}N_2$ to particulate N. The $^{15}N_2$ tracer is added into the ambient pool of N_2, the $^{15}N/^{14}N$ ratio is measured in the particulate N after incubation and compared to the natural abundance of N isotopes in unlabeled particulate material. The C_2H_2 method estimates N_2 fixation rate indirectly by measuring the reduction of C_2H_2 (a competitive inhibitor of N_2) to ethylene (C_2H_4) which is then converted to a N_2 fixation rate assuming of 3 or 4 moles C_2H_2 reduced per 1 mole N_2 fixed, depending on the extent of nitrogenase linked hydrogen production (Postgate, 1998). Generally, the direct $^{15}N_2$ assimilation is a precise and sensitive method; hence, it has been used to generate the majority of rates (Table 1b).

ESSDD
5, 47–106, 2012

Direct comparison of these two methods showed the direct $^{15}N_2$ assimilation method generally yields lower rates than those estimated from C_2H_2 reduction assay (see summary in Mulholland, 2007). Discrepancies between these measures could be because the $^{15}N_2$ assimilation method measures the net rate of conversion of reduced N to cellular N, or net N_2 fixation, while the C_2H_2 reduction method measures gross N_2 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 $^{15}N_2$ assimilation method significantly underestimates the N_2 fixation rates because the $^{15}N_2$ bubbles injected in seawater do not attain equilibrium with surrounding water (Mohr et al., 2010). However, there is no significant difference between our collected depth-integrated and volumetric, whole seawater N_2 fixation rates by the $^{15}N_2$ assimilation method ($3.3 \pm 18\,\mu mol\,N\,m^{-3}\,d^{-1}$, $n = 1950$) and those by the C_2H_2 reduction assays ($2.1 \pm 4.1\,\mu mol\,N\,m^{-3}\,d^{-1}$, $n = 125$) (two-sample Student's t-test, $p = 0.48$) (Fig. 1). Thus data acquired by either method are included in the database. 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_2 fixation rate measurement. Users have to be careful when using the database to study N_2 fixation rates aggregated from two different methods.

The collected N_2 fixation rates were mostly measured for whole seawater samples (Table 1b). Some samples were filtered and N_2 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 heterocystous cyanobacteria may be included in <10 µm 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,

ESSDD
5, 47–106, 2012

2008 and references therein). Thus those hourly N_2 fixation rates were converted to daily rates by multiplying with 12 h, which, however, could be conservative if the 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 qPCR detection limit varies according to several factors specific to each lab including sample volume, final extraction volume, and the volume of the DNA extract used in the qPCR reaction, but is typically in the range of 6–12 *nifH* copies l^{-1}. The limit of quantitation is significantly higher, typically on the order of 50–100 *nifH* copies l^{-1}. Thus we have assigned a value of 10 *nifH* copies l^{-1} to those data points reported as "detected but not quantifiable".

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,

ESSDD
5, 47–106, 2012

not just the heterocysts, of the heterocystous cyanobacteria. Limitations are associated with this extrapolation. Evidence indicates that this extrapolation overestimate 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 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, and are often not normally distributed and are positively skewed (long tail of high values). However, the datasets (excluding 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

ESSDD
5, 47–106, 2012

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).

We control for data quality by using Chauvenet's criterion to remove 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 remove 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 σ_{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 removed.

ESSDD
5, 47–106, 2012

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^{-1} and 100 cells trichome^{-1}. 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) (Table 3). 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 pg and V is the cell volume in μm^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 pg C cell^{-1} (Table 3) or 6.5 ng C trichome^{-1} (using 100 cells trichome^{-1}). The estimated carbon content for other *Trichodesmium* species is 110–250 pg C cell^{-1} (Table 3) or 11–25 ng C trichome^{-1} (using 100 cells trichome^{-1}). 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 ± 12 ng C trichome^{-1}. Some other studies observed higher C contents for *Trichodesmium* colonies, such as 9.7 µg C colony^{-1} in the Pacific (Mague et al., 1977) and 10.9 and 11.6 µg C colony^{-1} for *Trichodesmium* puffs and tufts in Atlantic (McCarthy and Carpenter, 1979), which results in a conversion factor ~50 ng C trichome^{-1} by assuming 200 trichomes colony^{-1}. Another direct analysis of *Trichodesmium* C content at the BATS station in the Sargasso Sea in 1995–1997 shows 8.3 ± 3.5 µg C colony^{-1} (n = 23) for *Trichodesmium* puffs, 9.9 ± 2.5 µg C colony^{-1} (n = 14) for *Trichodesmium* tufts and 4.5 ± 2.4 µg C colony^{-1} (n = 6) for *T. erythraeum* (Orcutt and Gundersen, unpublished data), which, by assuming 200 trichomes colony^{-1},

ESSDD
5, 47–106, 2012

leads to a conversion factor of 41, 49 and 22 ng C trichome^{-1}, respectively. To accommodate all these estimates, we use conversion factors of 30 ng C trichome^{-1} and 300 pg C cell^{-1} (using 100 cells trichome^{-1}) for *Trichodesmium* (Table 2). The latter is used for *nifH*-based *Trichodesmium* abundances as they are reported as number of cells.

UCYN-A cells are spherical in shape with a diameter of about 1 µm (Goebel et al., 2008), which gives a cell size estimate of 0.5 µm^3 and a carbon content of about 0.2 pg C cell^{-1} using the Verity et al. (1992) model (Table 2). UCYN-B (*Crocosphaera*) cells are also spherical in shape with a reported diameter range of 3–5 µm in laboratory isolates (Goebel et al., 2008) and 3–8 µm in natural samples and cultures (Webb et al., 2009; Moisander et al., 2010). Thus UCYN-B cells have a range in volume from 14–270 µm^3 and cellular carbon content between 4–54 pg C cell^{-1} (Verity et al., 1992). By assuming a diameter of 5 µm and thus a volume of 65 µm^3, we calculate a conversion factor of 20 pg C cell^{-1} for UCYN-B (Table 2). The only successful isolation and laboratory culture of a UCYN-C strain, designated TW3 (Taniuchi et al., 2011), show that the cells are 2.5–3.0 µm in width and 4.0–6.0 µm in length, which gives cellular volume of 30 µm^3 by using the middle values of the ranges and leads to a conversion factor of 10 pg C cell^{-1} for UCYN-C by using the Verity et al. (1992) model (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 (Table 4). As the sizes and the C contents are different in vegetative cells and heterocysts, average C content per *Richelia* or *Calothrix* cell are 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), and a biomass conversion factor of 10 pg C cell^{-1} is used for both *Richelia* and *Calothrix* (Table 2).

ESSDD
5, 47–106, 2012

The uncertainties related to these conversion factors will be discussed below.

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 S1h). By applying Chauvenet's criterion, there are only 8 data points rejected as their values are higher than the critical values, including 1 data point of volumetric *Trichodesmium* cell count which are greater than the geometric mean by 6 orders of magnitude (Fig. 2a), 1 data point of volumetric *Trichodesmium* N_2 fixation rate which are greater than the geometric mean by 5 orders of magnitude (Fig. 2b), 2 data points of volumetric N_2 fixation rate by UCYN which are greater than the geometric mean by 3 orders of magnitude (Fig. 2c), 1 data point of volumetric whole seawater N_2 fixation rate which are greater than the geometric mean by 3 orders of magnitude (Fig. 2d) and 3 data points of volumetric UCYN-B *nifH* genes which are greater than the geometric mean by 5 orders of magnitude (Fig. 2e). Thus from a statistical point of view, most data are acceptable. The underlying assumption of applying Chauvenet's criterion is that the samples are from similar biogeographic domains. The rejected 1 data point of *Trichodesmium* cell count of 4.4×10^{10} trichomes m^{-3} and 1 data point of *Trichodesmium* N_2 fixation rate of 31 391 µmol N m^{-3} 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_2 fixation rate were observed, these two data points were very likely real. The rejected data points of UCYN's N_2 fixation rate of 360 and 960 µmol N m^{-3} d^{-1} (Montoya et al., 2004) are from ~10° S, ~130–135° E, the western tropical Pacific in the Arafura Sea (Fig. 3b).

ESSDD
5, 47–106, 2012

High N_2 fixation rates are commonly found in this region, such as another rejected whole seawater N_2 fixation rate of 610 µmol N m^{-3} d^{-1} (Bonnet et al., 2009) found in the nearby sea region at 6° S, 147° E, near Papua New Guinea (Fig. 3b). Recent measurements in the southwestern Pacific near New Caledonia also show extremely high N_2 fixation rates (Bonnet, unpublished data). The possible reason for these high N_2 fixation rates could be the iron supply from volcanoes or via upwelling. The rejected 3 data points of UCYN *nifH*-based abundances of ~2–9 × 10^{11} copies m^{-3} (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 3 rejected data points of the extreme peaks in *nifH*-based abundance were sampled in June and July 2009 during extremely high water temperature events (~30 °C) towards the end of the summer. This is also a time of the year when the lowest seasonal dissolved inorganic N:P-ratios (approximately 0.5) occurred, although the ratio is increasing 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 Chl-*a* (5–6 µg l^{-1}). Thus we believe the rejected data points 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 that in the Arabian

ESSDD
5, 47–106, 2012

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 continuous time-series of N_2 fixation measurements.

Most data were collected in the tropical and subtropical regions, with latitudinal coverage of 50° S–50° N for cell counts of diazotrophs, 40° S–60° N for N_2 fixation rates and 30° S–60° 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 spring and early summer (Fig. 5f).

3.3 N_2 fixation rates

After being binned onto $3 \times 3^\circ$ 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 $\mu mol\, N\, m^{-2}\, d^{-1}$ (Fig. 6a). In most other regions, depth-integrated N_2 fixation rates are on an order of 1–100 $\mu mol\, N\, m^{-2}\, d^{-1}$ (Fig. 6a).

The volumetric N_2 fixation rates are also analyzed on $3 \times 3^\circ$ grids in the five vertical layers of 0–5, 5–25, 25–62.5, 62.5–137.5 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

ESSDD
5, 47–106, 2012

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).

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 contributors measured the conversion factors. The geometric mean of cell counts in both the depth-integrated and volumetric datasets demonstrates that *Trichodesmium* is the dominant diazotrophs with abundances 1 order of magnitude higher than heterocystous cyanobacteria (*Richelia* and *Calothrix*), except that the abundance of depth-integrated *Calothrix* is in same order as *Trichodesmium* (Table 5). But 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 data). As the average *Trichodesmium* cell size, and thus the biomass conversion 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 show 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

ESSDD
5, 47–106, 2012

trichome^{-1}) and *nifH*-based data are in agreement (Fig. 7). This evidence provides some support for the reliability of *nifH*-based abundance for *Trichodesmium* and our assumption of converting *nifH* gene copies to cell numbers on 1:1 basis.

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, 5–25, 25–62.5, 62.5–137.5 and 137.5–250 m by showing the geometric means of each $3 \times 3^{\circ}$ 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 hetercystous 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 to 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 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.

ESSDD
5, 47–106, 2012

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, 5–25, 25–62.5, 62.5–137.5 and 137.5–250 m by showing the geometric means of each 3 × 3° 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 ~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, Mediterranean Sea, the North and the South Pacific Ocean and the Indian Ocean (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 × 3° 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 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 ~40° S–55° 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,

ESSDD
5, 47–106, 2012

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 no depth-integrated N_2 fixation rate collected in the Indian Ocean. By summing the geometric mean rates from all the regions, the global N_2 fixation rate (not including the Indian Ocean) is estimated to 62 (error range: 53–73) Tg N yr^{-1} (Table 6), which is consistent with 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). 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

ESSDD
5, 47–106, 2012

counted (Table 1a) and thus very low total diazotrophic biomass was reported for these sites (Fig. 8a), we take a special handling 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 higher than that in the North Pacific as shown by both geometric and arithmetic means (Table 7), which is inconsistent with the pattern of N_2 fixation rates. This finding could indicate that specific N_2 fixation rates are higher in the North Pacific than in the North Atlantic, or it could simply be a result of undersampling in the vast North Pacific. The global diazotrophic biomass is estimated from the cell-count-based data as 4.7 (error range: 2.3–9.6) Tg C using geometric mean or 33 (standard error: 8.4) Tg C using arithmetic mean (Table 7).

The global diazotrophic biomass is also estimated from depth-integrated *nifH*-based data, which are available in the North Atlantic Ocean, the North and the South Pacific Ocean and the Indian Ocean. First, the depth-integrated data points of *nifH*-based biomass are limited (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 (not including the South Atlantic) of diazotrophic biomass from *nifH*-based data is 89 (error range: 40–200) Tg C using geometric mean or 640 (standard error: 150) 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 due to N_2 fixation is estimated to be 5 days and 92 days, respectively.

However, there are a number of limitations with these estimates. Foremost, these data points are not uniformly distributed in the world's 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

ESSDD
5, 47–106, 2012

relative to true regional means. As the current database does not cover some regions such as the coastal upwelling zones, our estimate could be changed substantially if values differ within these biogeographic domains. The log-normal distributions also reveal high variance of the measurements. As indicated by the geometric error range (Tables 6, 7 and 8), high uncertainties of the global estimated mean still exist especially for the diazotrophic biomass. Based on depth-integrated data, the global mean estimate of diazotrophic biomass from cell counts is 1 order of magnitude lower than that from *nifH*-based abundances (Tables 7 and 8), which are mostly contributed by the dominated taxonomy, *Trichodesmium* (Table 5). But considering the log-normal distributions, difference of one order of magnitude is relatively small compared to the span of the distribution especially when 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.

3.6 Uncertainties of biomass conversion factors

As described above, the default biomass conversion factor for *Trichodesmium* was selected by merging estimates from cell sizes of cultured *Trichodesmium* with direct carbon content measurements for *Trichodesmium* colonies. A compilation of the available data has shown that ~100 cells per trichome was widely used as an estimate for *Trichodesmium* (LaRoche and Breitbarth, 2005). Thus we do not consider the variations for the conversion factors between *Trichodesmium* cells and trichomes, focusing instead on the variations of *Trichodesmium* cell carbon content. *T. erythraeum* carbon content was back-calculated to 42 ± 1 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), which is consistent with our estimates from cell sizes for this species (Table 3). Thus the cell size-carbon content model we used, the Verity et al. (1992) model, appears to be suitable for *Trichodesmium*. However, *T. erythraeum* is relatively small compared to other *Trichodesmium* species. The carbon

ESSDD
5, 47–106, 2012

content for other *Trichodesmium* species is at least 110 pg C cell^{-1} (Table 3). Also, direct elemental analysis for *Trichodesmium* by Carpenter et al. (2004) report values of 10 000 ± 12 000 pg C trichome^{-1} or 100 ± 120 pg C cell^{-1} (100 cells trichome^{-1} assumed). Of course, the overall conversion factor depends on the composition of *Trichodesmium* species, which is unknown and could be variable, temporally and spatially (Hynes, 2009). Thus we estimate 100 pg C cell^{-1} and 10 000 pg C trichome^{-1} as the low end for the conversion factor for *Trichodesmium* (Table 9). Direct measurements made at the BATS station yield the conversion factor of 450 pg C cell^{-1} as described above. The synthesis of laboratory and field data by LaRoche and Breitbarth (2005) provides values of 50 000–58 000 pg C trichome^{-1}, which equals ~500 pg C cell^{-1} when assuming 100 cells trichome^{-1}. These lines of evidence indicate that the *Trichodesmium* cell size in the field may be larger than that of cultured species. Thus, we estimate 500 pg C cell^{-1} and 50 000 pg C trichome^{-1} (assuming 100 cells trichome^{-1}) as the upper limits for the *Trichodesmium*-specific biomass conversion factors (Table 9).

The biomass of UCYN-A is difficult to calculate because there is no isolate in culture. The estimate of ~1 µm 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. Thus a range for the biomass conversion factor is estimated for UCYN-A by varying the default conversion factor by ±50 %, i.e. 0.1–0.3 pg C cell^{-1} (Table 9). Also, as discussed above, the size of UCYN-B ranges from 3–8 µm in diameter, which leads to a conversion factor range of 4–50 pg C cell^{-1} by using the Verity et al. (1992) model (Table 9). The size range of the isolated UCYN-C strain of 2.5–3.0 µm in width and 4.0–6.0 µm in length (Taniuchi et al., 2011) results in a conversion factor of 5–9 pg C cell^{-1} 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 4–5 µm width by 7–8 µm length (Reddy et al., 1993), which equals 90–160 µm^3 and leads to a conversion factor of 15–24 pg C cell^{-1} by using the Verity et al. (1992) model.

ESSDD
5, 47–106, 2012

By merging these two estimates, a conversion factor range of 5–24 pg C cell^{-1} is used for UCYN-C (Table 9).

Based on the estimates made in Table 4, the ranges of the biomass conversion factors of 2–80 pg C cell^{-1} and 5–20 pg C cell^{-1} are used for *Richelia* and *Calothrix*, respectively (Table 9).

In order to show the effects of the variation of the biomass conversion factors in a more straightforward manner, 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 in the previous section. The estimated global geometric mean diazotrophic biomass can vary by about ±70 % from default estimates, ranging in 1.6–7.7 Tg C 9 (default 4.7 Tg C) based on cell count data, or in 27–170 Tg C (default 89 Tg C) based on *nifH*-based data (Table 9).

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 are another concern. Although careful analyses have been conducted to derive a reliable global biomass, our estimates may vary by 5–6 fold 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

ESSDD
5, 47–106, 2012

data. The estimated global mean diazotrophic biomass from *nifH*-based abundances is 1 order 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.

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.

Supplementary material related to this article is available online at: http://www.earth-syst-sci-data-discuss.net/5/47/2012/essdd-5-47-2012-supplement.pdf.

Acknowledgements. Y.-W. L., S. C. D., M. J. C. and D. M. K. 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. and R. M. L. 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. J. P. Z., P. H. M. and K. A. T. are partially supported by C-MORE and 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.

ESSDD
5, 47–106, 2012

A. P. R. received funding through NERC OCEANS2025. A. E. W. acknowledges funding from NSF OCE-0726422 (to F. Prahl, OSU).

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.

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 (BIOSOPE cruise), Biogeosciences, 5, 215–225, doi:10.5194/bg-5-215-2008, 2008.

Bonnet, S., Biegala, I. C., Dutrieux, P., Slemons, 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 occurence 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

ESSDD
5, 47–106, 2012

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., Bednarsek, 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.: Introduction: MAREDAT – towards a world atlas of marine plankton functional types, 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., 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_2-fixing cyanobacterium *Trichodesmium erythraeum* in the central Arabian Sea, Mar. Ecol. Prog. Ser., 172, 281–292, doi:10.3354/meps172281, 1998.
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., 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,

ESSDD
5, 47–106, 2012

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_2 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. Microb., 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

ESSDD
5, 47–106, 2012

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_2 fixation by unicellular bacterioplankton from the Atlantic and Pacific oceans: Phylogeny and in situ rates, Appl. Environ. Microb., 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_2 fixation in the Atlantic Ocean, Biogeosciences, 7, 3167–3176, doi:10.5194/bg-7-3167-2010, 2010.
Fernández, A., Graña, R., Mouriño-Carballido, B., Bode, A., Varela, M., Domínguez, J. F., Escánez, J., de Armas, D., and Marañón, E.: Community N_2 fixation and *Trichodesmium* spp. abundance along longitudinal gradients in the Eastern Subtropical North Atlantic, Earth Syst. Sci. Data Discuss., in preparation, 2012.
Fernández, C., Farias, L., and Ulloa, O.: Nitrogen Fixation in Denitrified Marine Waters, Plos One, 6, e20539, doi:10.1371/journal.pone.0020539, 2011.
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

ESSDD
5, 47–106, 2012

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. Microbiol., 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_2 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.
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_2 fixation in the tropical Atlantic Ocean,

ESSDD
5, 47–106, 2012

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.
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.
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 thesSouthewestern Sagasso 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.: Diversity of the marine cyanobacterium *Trichodesmium*: characterization of the Woods Hole culture collection and quantification of field populations, Ph.D., MIT-WHOI Joint Program, Massachusetts Institute of Technology, Cambridge, MA, 162 pp., 2009.
Hynes, A. M., Webb, E. A., Doney, S. C., and Waterbury, J. B.: Comparison of cultured

ESSDD
5, 47–106, 2012

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. Microb., 71, 7910–7919, doi:10.1128/Aem.71.12.7910-7919.2005, 2005.

Langlois, R. J., Hümmer, D., and LaRoche, J.: Abundances and distributions of the dominant *nifH* phylotypes in the Northern Atlantic Ocean, Appl. Environ. Microb., 74, 1922–1931, doi:10.1128/Aem.01720-07, 2008.

LaRoche, J. and Breitbarth, 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.

ESSDD

5, 47–106, 2012

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. Microb., 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_2 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_2 fixation, Appl. Environ. Microb., 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_2 fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean, Nature, 430,

ESSDD
5, 47–106, 2012

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_2 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 and Zehr, J. P.: Rates of dinitrogen fixation and the abundance of diazotrophs in North American coastal waters between Cape Hatteras and Georges Bank, in preparation, 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_2 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.
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.

ESSDD
5, 47–106, 2012

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, *Osciflutoriu thiebuutii*, 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_2 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* Ostenfeld (Bacillariophyceae) and *Richelia intracellularis* Schmidt (Cyanophyceae), Bot. Mar., 27, 345–355, 1984.
Taniuchi, Y., Chen, Y.-l. 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

ESSDD
5, 47–106, 2012

Ocean, Environ. Microbiol., published online, doi:10.1111/j.1462-2920.2011.02606.x, 2011.
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, Kluwer Academic Publishers, Boston, 362, 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,

ESSDD
5, 47–106, 2012

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, Crocosphaera 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., Verdeny, 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, 2011.
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.
Zeev, E. B., 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.
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. Microb., 64, 3444–3450, 1998.
Zehr, J. P., Waterbury, J. B., Turner, P. J., Montoya, J. P., Omoregie, E., Steward, G. F., Hansen, A., and Karl, D. M.: Unicellular cyanobacteria fix N_2 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., Niazi, F., Shi, T., Tripp, H. J., and Affourtit, J. P.: Globally Distributed Uncultivated Oceanic N_2-Fixing Cyanobacteria Lack Oxygenic Photosystem II, Science, 322, 1110–1112, doi:10.1126/science.1165340, 2008.
Zehr, J. P.: Nitrogen fixation by marine cyanobacteria, Trends Microbiol., 19, 162–173, doi:10.1016/j.tim.2010.12.004, 2011.

ESSDD
5, 47–106, 2012

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
	Trichodesmium	Unicellular	Heterocystous	Depth Integral[a]	Total Points[c]		
Benavides				29[b]	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[b]	94	Subtropical N. Pacific	Sohm et al. (2011a)
Capone/Carpenter/Subramaniam	402		176	101[b]+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[b]	146	East China Sea	Chang et al. (2000)
Chen				47[b]	47	South China Sea	Chen et al. (2003)
Chen	20				20	upstream Kuroshio and South China Sea	Chen et al. (2008)
Chen				35[b]	35	Northwest Pacific (Kuroshio)	Chen et al. (2011)
Dong				36[b]	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, 2012); Mouriño-Carballido et al. (2011)
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
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[b]	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[b] +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)
Zeev			26		26	Mediterranean Sea	Zeev et al. (2008)
Total	3139	0	1612	930	5191		

[a] Computed from vertical profiles unless marked for those reported by data providers as depth-integrals.

[b] Data are reported by data providers as depth-integrated biomass.

[c] Do not include depth-integrated profiles unless they are reported by data providers as depth-integrated values (as marked by [b]).

ESSDD
5, 47–106, 2012

Table 1b. Summary of data points for N_2 fixation rates, including volumetric measurements of *Trichodesmium*, unicellular cyanobacteria and heterocystous cyanobacteria and their depth-integrals.

Investigator	N_2 Fixation Rates							Region	References
	Tricho -desmium	Uni-cellular	Hetero-cystous	Whole Seawater	Depth Integral[d]	Total Points[f]	method		
Benavides		53		60	32[e]	145	C_2H_2 reduction and $^{15}N_2$ assimilation	Subtropical N. Atlantic	Benavides et al. (2011)
Bonnet		27		20	10	47	$^{15}N_2$ assimilation	Tropical Pacific	Bonnet et al. (2009)
Bonnet		80			9	80	$^{15}N_2$ assimilation	Mediterranean Sea	Bonnet et al. (2011)
Church/Karl				332	50	332	$^{15}N_2$ assimilation	Subtropical N. Pacific (ALOHA)	Church et al. (2009)
Carpenter	157				39	157	C_2H_2 reduction	Tropical/subtropical W. Atlantic	Carpenter and Price (1977)
Capone/Carpenter/Subramaniam				180	40[e]+33	220	C_2H_2 reduction and $^{15}N_2$ assimilation	Subtropical N. Pacific	Sohm et al. (2011a)
Capone/Carpenter/Subramaniam	410		185		119[e]+118	714	C_2H_2 reduction	Tropical/subtropical N. Atlantic	Capone et al. (2005)
Church				7		7	$^{15}N_2$ assimilation	Subtropical N. Pacific	Fong et al. (2008)
Dore		13		24	6	37	$^{15}N_2$ assimilation	Subtropical N. Pacific (ALOHA)	Dore et al. (2002)
Falcón		24			6	24	$^{15}N_2$ assimilation	Subtropical N. Pacific (ALOHA) and Subtropical N. Atlantic	Falcón et al. (2004)
Fernández/Mouriño -Carballido/Marañón				115	32	115	$^{15}N_2$ assimilation	Cross N.-S. Atlantic Mouriño-Carballido et al. (2011)	Fernández et al. (2010, 2012);
Foster				1		1	$^{15}N_2$ assimilation	Subtropical N. Pacific	Needoba et al. (2007)
Furuya/Kitajima		24		29		53	C_2H_2 reduction	Tropical/subtropical N. Pacific	Kitajima et al. (2009)
Furuya/Shiozaki				103	11	103	$^{15}N_2$ assimilation	Tropical/subtropical Pacific	Shiozaki et al. (2010)
Goering	40				4	40	$^{15}N_2$ assimilation	Tropical Atlantic	Goering et al. (1966)
Gunderson				24	2	24	C_2H_2 reduction	Subtropical N. Pacific	Gundersen et al. (1976)
Karl	12	12		12	9	36	$^{15}N_2$ assimilation	Subtropical N. Pacific (ALOHA)	Grabowski et al. (2008)
Kromkamp	15					15	C_2H_2 reduction	W. Indian Ocean (African Coast)	Kromkamp et al. (1997)
Mague			8		1	8	C_2H_2 reduction	Subtropical N. Pacific	Mague et al. (1974)
Mague				46	8	46	C_2H_2 reduction	Subtropical N. Pacific	Mague et al. (1977)
McCarthy/Carpenter	24				5	24	C_2H_2 reduction	Subtropical N. Atlantic	McCarthy and Carpenter (1979)
Montoya		16			1	16	$^{15}N_2$ assimilation	Subtropical N. Pacific	Montoya et al. (2004)
Montoya		5				5	$^{15}N_2$ assimilation	Arafura Sea (near Australia)	Montoya et al. (2004)
Moore				46	10	46	$^{15}N_2$ assimilation	Tropical/subtropical Atlantic	Moore et al. (2009)
Mulholland					30[e]	30	$^{15}N_2$ assimilation	Subtropical NW. Atlantic	Mulholland et al. (2012)
Mulholland				181	26	181	$^{15}N_2$ assimilation	Subtropical NW. Atlantic	Mulholland et al. (2012)
Mulholland				185	36	185	$^{15}N_2$ assimilation	Subtropical N. Atlantic	Mulholland et al. (2012)
Orcutt/Gundersen	31					31	$^{15}N_2$ assimilation	Subtropical N. Atlantic (BATS)	Orcutt et al. (2001)
Raimbault				158	24	158	$^{15}N_2$ assimilation	Tropical/subtropical S. Pacific	Raimbault and Garcia (2008)
Raimbault				197	33	197	$^{15}N_2$ assimilation	Subtropical S. Pacific	Garcia et al. (2007)
Raimbault				71	12	71	$^{15}N_2$ assimilation	Mediterranean Sea	Sandroni et al. (2007)
Rees				5		5	$^{15}N_2$ assimilation	Eastern Mediterranean Sea	Rees et al. (2006)
Rees				21	7	21	$^{15}N_2$ assimilation	Subtropical N. Atlantic (FeeP)	unpublished data
Rees				70		70	$^{15}N_2$ assimilation	Cross N.-S. Atlantic (AMT14&15)	unpublished data
Rees				2		2	$^{15}N_2$ assimilation	Western English Channel	Rees et al. (2009)
Rees				57	5	57	$^{15}N_2$ assimilation	Tropical E. Atlantic	Turk et al. (2011)
Turk/Zehr				11	2	11	$^{15}N_2$ assimilation	Tropical Atlantic	Goebel et al. (2010)
White/Letelier		11		12	4	23	$^{15}N_2$ assimilation	Subtropical N. Pacific	White et al. (unpublished data)
White/Letelier		10		43		53	$^{15}N_2$ assimilation	Subtropical N. Pacific	Watkins-Brandt et al. (2011)
White				134	19	134	$^{15}N_2$ assimilation	Gulf of CA and Tropical Pacific	White et al. (2011)
Zeev			12			12	C_2H_2 reduction	Mediterranean Sea	Zeev et al. (2008)
Total	689	275	205	2146	743	3536			

[d] Computed from vertical profiles unless marked for those reported by data providers as depth-integrals.
[e] Data are reported by data providers as depth-integrated N_2 fixation rates.
[f] Do not include integrated profiles unless they are reported by data providers as depth-integrated values (as marked by [e]).

ESSDD
5, 47–106, 2012

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	*nifH*-based abundances					Region	References
	Trichodesmium	Unicellular	Hetero-cystous	Depth Integral	Total Points[g]		
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[h]	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)
Total	759	1793	599	416	3201		

[g] Do not include integrated profiles unless they are reported by data providers as depth-integrated values (as marked by [h]).
[h] Data are reported by data providers as depth-integrated values.

ESSDD
5, 47–106, 2012

Table 2. Default biomass conversion factors.

Trichodesmium (pg C trichome^{-1})	30 000
Trichodesmium (pg C cell^{-1})	300
UCYN-A (pg C cell^{-1})	0.2
UCYN-B (pg C cell^{-1})	20
UCYN-C (pg C cell^{-1})	10
Heterocystous *Richelia* (pg C trichome^{-1})	10
Heterocystous *Calothrix* (pg C trichome^{-1})	10

ESSDD
5, 47–106, 2012

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

ESSDD
5, 47–106, 2012

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 ± standard deviation when applicable.

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

ESSDD
5, 47–106, 2012

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	*n*	*n* of zero-value data	Abundance (10^9 cells m^{-2})		Biomass (mg C m^{-2})	
			Geometric	Arithmetic	Geometric	Arithmetic
cell-count-based data						
*Trichodesmium**	620	59	0.16 (0.14–0.18)	1.2 (0.12)	43 (38–49)	310 (27)
Richelia	288	41	0.010 (0.008–0.013)	3.3 (0.88)	0.099 (0.078–0.13)	33 (8.8)
Calotrhix	17	9	0.084 (0.050–0.14)	0.096 (0.050)	0.84 (0.50–1.4)	0.96 (0.50)
nifH-based data						
Trichodesmium	103	10	1.4 (1.1–1.9)	8.8 (1.5)	400 (310–530)	2,600 (440)
UCYN-A	104	24	1.5 (0.88–2.4)	3,400 (2,200)	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)
Richelia	63	2	0.24 (0.15–0.38)	34 (12)	2.4 (1.5–3.9)	340 (120)
Calotrhix	11	0	0.21 (0.093–0.46)	0.58 (0.15)	2.1 (0.93–4.6)	5.8 (1.5)
Volumetric Data	*n*	*n* of zero-value data	Abundance (10^6 cells m^{-3})		Biomass (µg C m^{-3})	
			Geometric	Arithmetic	Geometric	Arithmetic
cell-count-based data						
*Trichodesmium**	3087	1019	3.3 (3.1–3.5)	30 (3.8)	1100 (990–1100)	9700 (1200)
Richelia	1564	781	0.11 (0.097–0.12)	11 (1.5)	1.1 (0.97–1.2)	110 (15)
Calotrhix	48	28	0.091 (0.060–0.14)	0.13 (0.043)	0.91 (0.60–1.4)	1.3 (0.43)
nifH-based data						
Trichodesmium	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)
Richelia	535	160	1.6 (1.3–1.9)	240 (63)	16 (13–19)	2400 (630)
Calothrix	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^{-1}.

ESSDD
5, 47–106, 2012

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 $3 \times 3^{\circ}$ 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.

Region	Latitudinal range	Ocean area ($\times 10^{12}$ m^2)	*n* of bins	*n* of zero-value bins	N_2 fixation rate (μmol N m^{-2} d^{-1})		Areal Sum (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	16	0	7.6 (6.1–9.3)	12 (4.1)	1.1 (0.86–1.3)	1.8 (0.58)
Mediterranean Sea	30–45° N	6.3	7	0	21 (11–41)	60 (29)	0.67 (0.34–1.3)	1.9 (0.92)
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	ND	ND	ND	ND	ND	ND
Global							62 (*53–73*)	140 (9.2)

ESSDD
5, 47–106, 2012

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.

Region	Latitudinal range	Ocean area ($\times 10^{12}$ m^2)	*n* of bins	*n* of zero-value bins	Cell-count-based diazotrophic biomass (mg C m^{-2})		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	9	4	13 (8.4–19)	9.2 (3.8)	0.35 (0.23–0.53)	0.25 (0.10)
Mediterranean Sea	30–45° N	6.3	ND	ND	ND	ND	ND	ND
North Pacific Ocean	0–55° N	89						
Trichodesmium			48	15	10 (6.3–17)	78 (21)		
Heterocystous CYN			61	11	0.03 (0.02–0.04)	0.44 (0.12)		
Total					10 (6.3–17)	78 (16)	0.93 (0.56–1.53)	6.92 (1.5)
South Pacific Ocean	40–0° S	72						
Trichodesmium			4	1	0.11 (0.09–0.13)	0.09 (0.03)		
Heterocystous CYN			25	20	0.04 (0.03–0.06)	0.012 (0.006)		
Total					0.15 (0.12–0.20)	0.10 (0.01)	0.011 (0.008–0.014)	0.007 (0.001)
Indian Ocean	40° S–25° N	62	3	0	41 (8.5–190)	240 (130)	2.5 (0.52–12)	15 (7.8)
Global							4.7 (*2.0–15*)	33 (2.0)

ESSDD
5, 47–106, 2012

Table 8. Estimates of diazotrophic biomass for the global oceans using *nifH*-based data. Same method as Table 6 is used. ND: no data.

Region	Latitudinal range	Ocean area ($\times 10^{12}$ m^2)	*n* of bins	*n* of zero-value bins	*nifH*-based diazotrophic biomass (mg C m^{-2})		Areal Sum (Tg C)	
					Geometric	Arithmetic	Geometric	Arithmetic
North Atlantic Ocean	0–55° N	36	39	6	310 (160–620)	3900 (1300)	11 (5.7–22)	140 (46)
South Atlantic Ocean	40–0 °S	27	ND	ND	ND	ND	ND	ND
Mediterranean Sea	30–45° N	6.3	ND	ND	ND	ND	ND	ND
North Pacific Ocean	0–55° N	89	6	0	68 (15–300)	1800 (1100)	6.0 (1.3–27)	160 (96)
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	6	0	140 (71–270)	510 (210)	8.5 (4.4–16)	31 (13)
Global							89 (*52–170*)	640 (73)

ESSDD
5, 47–106, 2012

Table 9. 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^{-1})			Global Biomass Estimate[b] (Tg C)		
	default	lower	upper	default	lower	upper
Cell-count-based biomass						
Trichodesmium[a]	300	100	500			
Richelia	10	2	80			
Calothrix	10	5	20			
				4.7	1.6	7.7
nifH-based biomass						
Trichodesmium	300	100	500			
UCYN-A	0.2	0.1	0.3			
UCYN-B	20	4	50			
UCYN-C	10	5	24			
UCYN all groups[b]						
Richelia	10	2	80			
Calothrix	10	5	20			
				89	27	170

[a] Assuming 100 cells trichome^{-1}.

[b] The low/high biomass are calculated when all their subtypes use low/high conversion factors.

ESSDD
5, 47–106, 2012

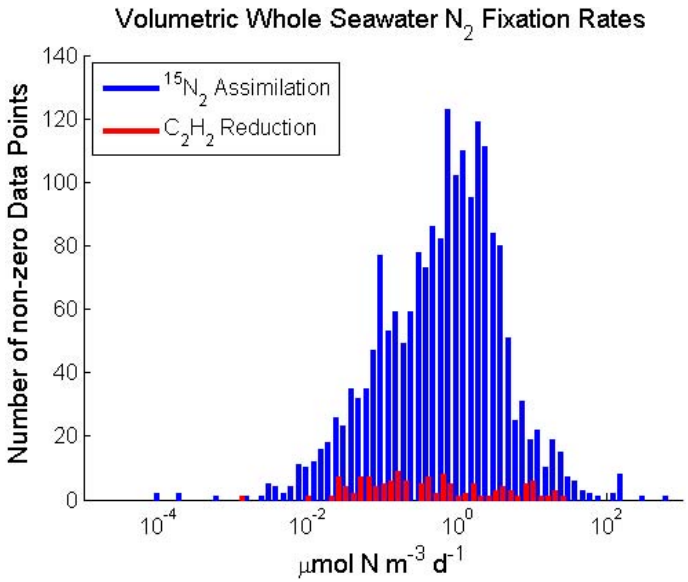

Fig. 1. Histogram of volumetric, whole seawater N_2 fixation rates (non-zero values) from $^{15}N_2$ assimilation assays and C_2H_2 reduction assays. Data values are on logarithmic scale.

ESSDD
5, 47–106, 2012

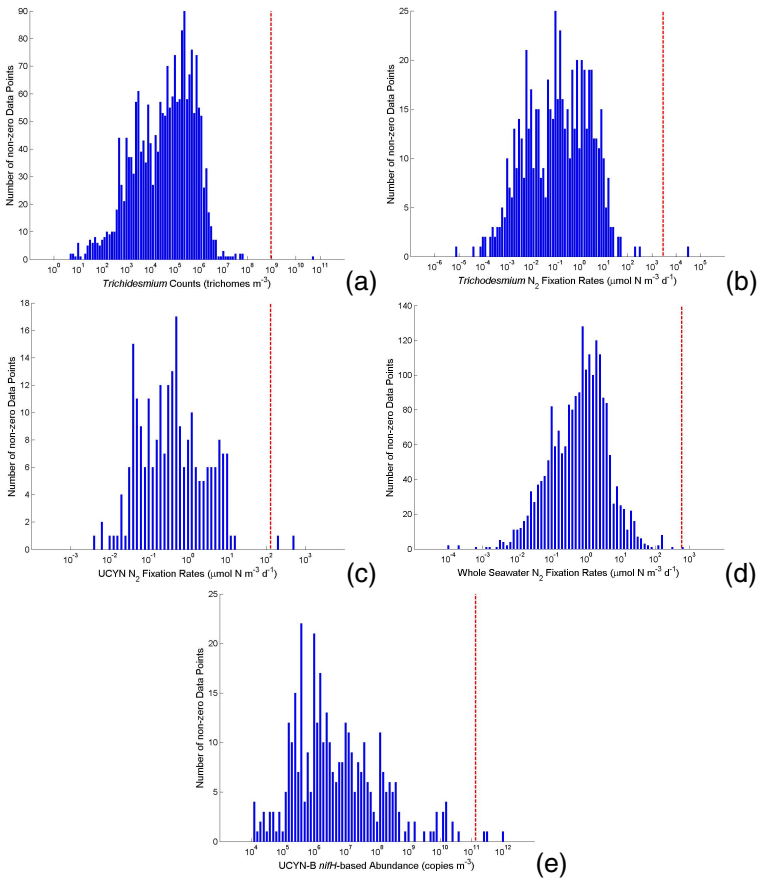

Fig. 2. Histogram of data points on logarithmic scale (blue bars) and the critical values for quality control using Chauvenet's criterion (dashed red lines). Values higher than the critical values are rejected. **(a)** *Trichodesmium* cell counts, **(b)** *Trichodesmium* N_2 fixation rates, **(c)** UCYN N_2 fixation rates, **(d)** whole seawater N_2 fixation rates, and **(e)** UCYN-B *nifH*-based abundance. See Supplement Fig. S1 for figures for other types.

ESSDD
5, 47–106, 2012

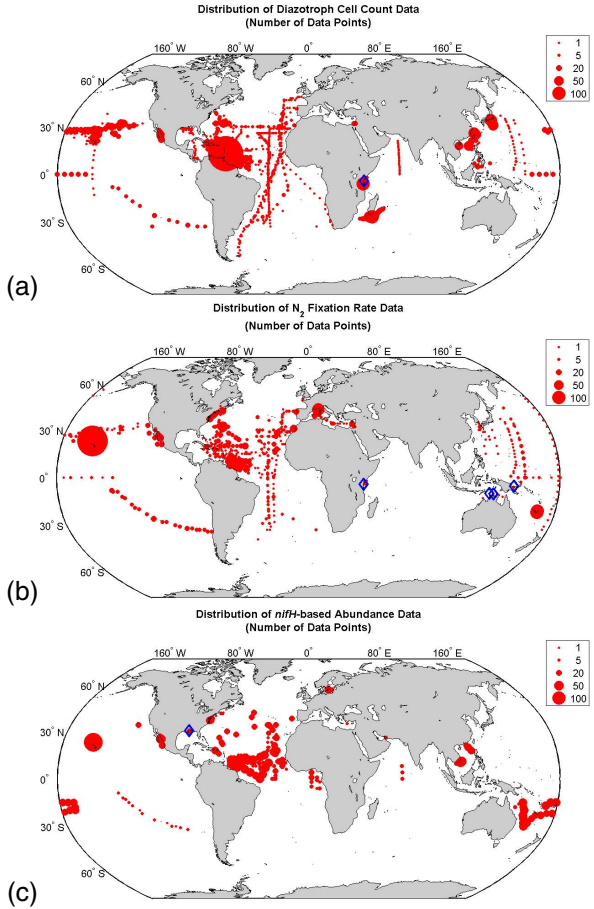

Fig. 3. Spatial distributions of collected diazotrophic data in number of data points (binned on to $1 \times 1^{\circ}$ grid), including cell counts (panel **a**), N_2 fixation rates (panel **b**) and *nifH*-based abundances (panel **c**). Blue diamonds mark the location for the data rejected by Chauvenet's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

ESSDD
5, 47–106, 2012

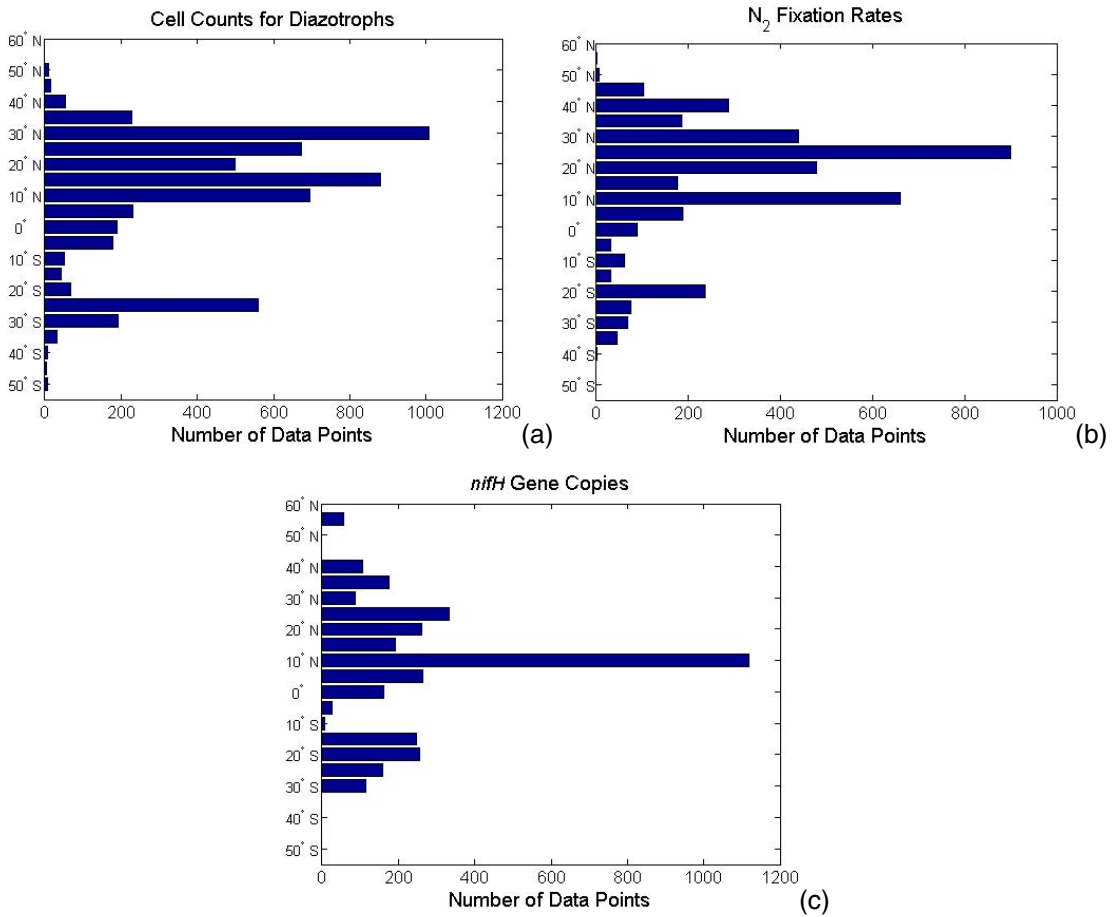

Fig. 4. Latitudinal distribution of the data for **(a)** cell counts for diazotrophs, **(b)** N_2 fixation rates, and **(c)** *nifH*-based abundances.

ESSDD
5, 47–106, 2012

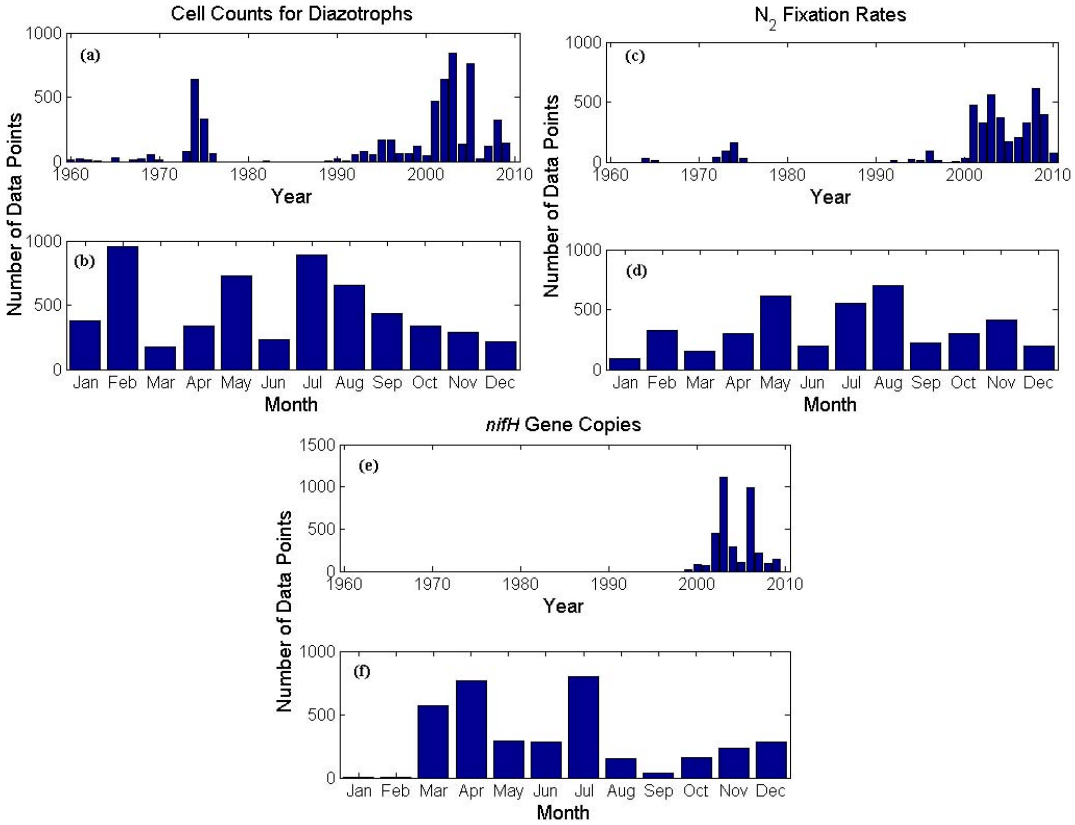

Fig. 5. Temporal distribution of the data points in year and month for **(a)**–**(b)** cell counts for diazotrophs, **(c)**–**(d)** N_2 fixation rates, and **(e)**–**(f)** *nifH*-based abundances.

ESSDD
5, 47–106, 2012

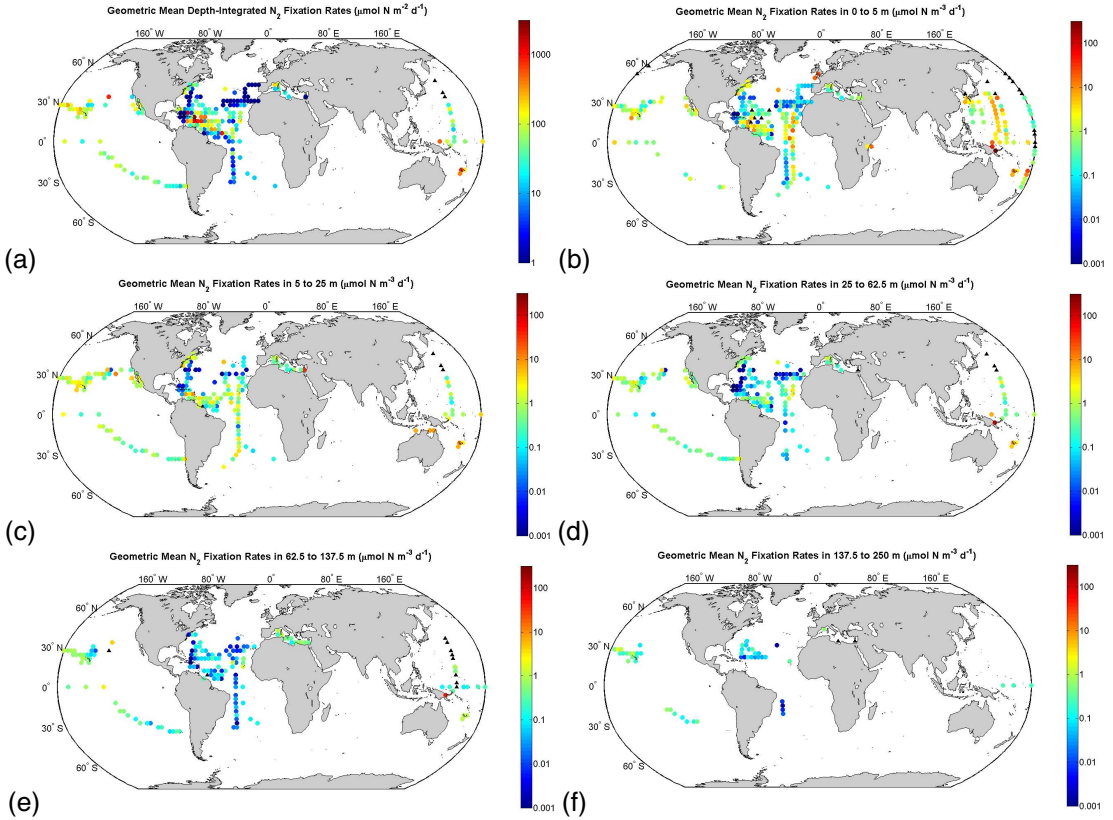

Fig. 6. **(a)** Geometric mean depth-integrated N_2 fixation rates binned on $3 \times 3^{\circ}$ grid. Also geometric mean N_2 fixation rates on $3 \times 3^{\circ}$ grid for 5 vertical layers of **(b)** 0–5 m, **(c)** 5–25 m, **(d)** 25–62.5 m, **(e)** 62.5–137.5 m and **(f)** 137.5–250 m. The color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

ESSDD
5, 47–106, 2012

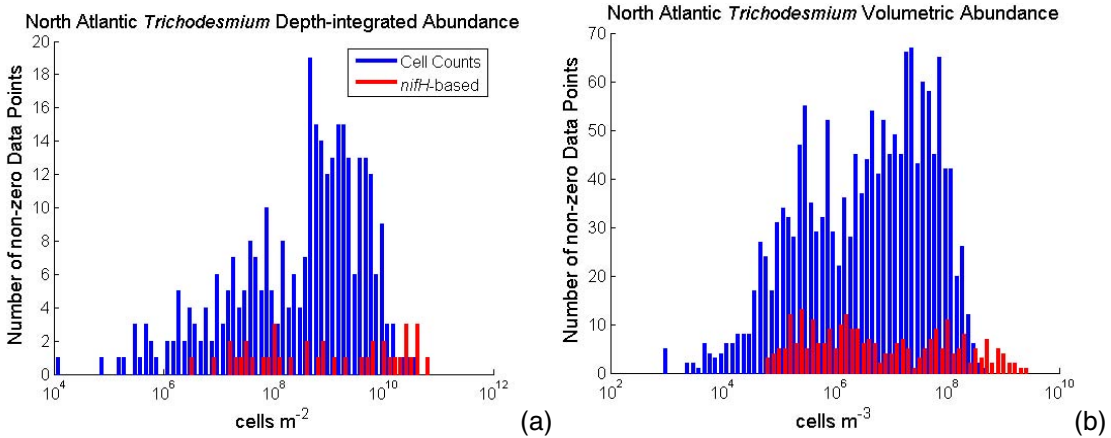

Fig. 7. Histogram of depth-integrated (panel **a**) and volumetric (panel **b**) non-zero data points of cell-count-based (blue) and *nifH*-based (red) *Trichodesmium* abundances in the North Atlantic (10° S–50° N). Each *Trichodesmium* trichome is assumed to be comprised of 100 cells. Data values are on logarithmic scale. Those "detected but not quantifiable" *nifH*-based abundances, which are assigned 5×10^3 cells m^{-3} (see text for details), are not included.

ESSDD
5, 47–106, 2012

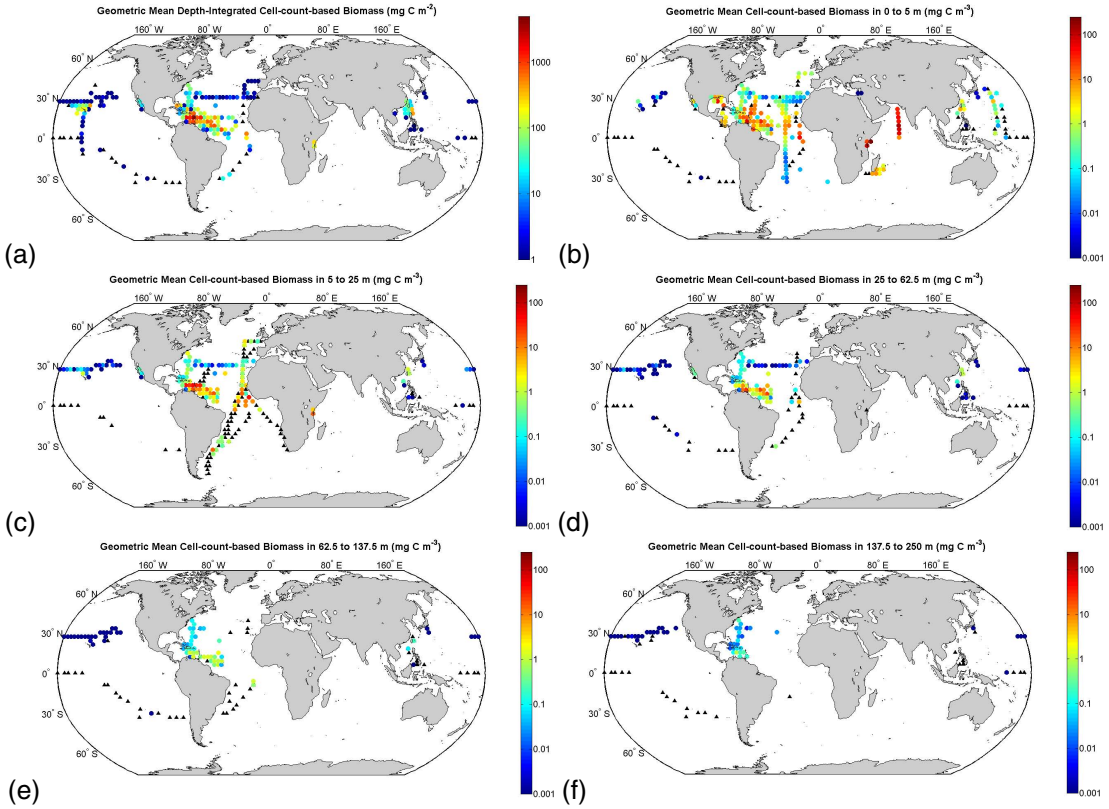

Fig. 8. **(a)** Geometric mean depth-integrated cell-count-based diazotrophic carbon biomass binned on $3 \times 3^{\circ}$ grid. Also geometric mean cell-count-based diazotrophic carbon biomass on $3 \times 3^{\circ}$ grid for 5 vertical layers of **(b)** 0–5 m, **(c)** 5–25 m, **(d)** 25–62.5 m, **(e)** 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.

ESSDD
5, 47–106, 2012

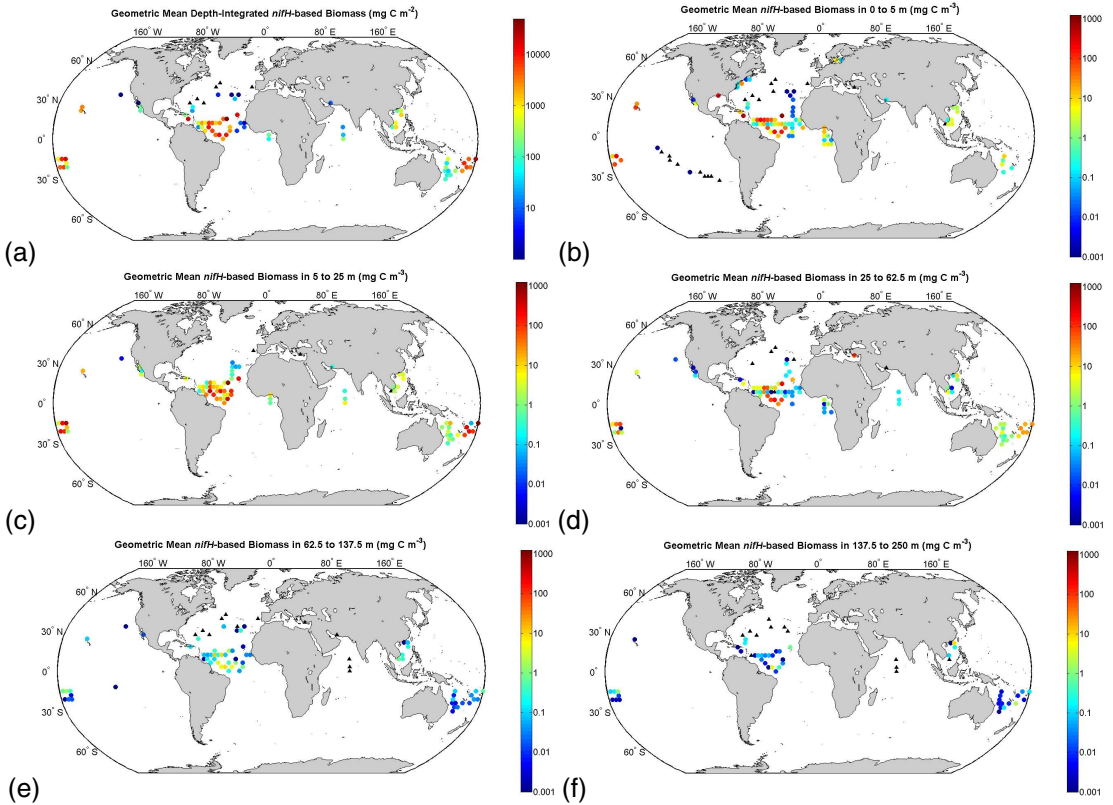

Fig. 9. **(a)** Geometric mean depth-integrated *nifH*-based diazotrophic carbon biomass binned on 3 × 3° grid. Also geometric mean *nifH*-based diazotrophic carbon biomass on 3 × 3° grid for 5 vertical layers of **(b)** 0–5 m, **(c)** 5–25 m, **(d)** 25–62.5 m, **(e)** 62.5–137.5 m and **(f)** 137.5–250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

ESSDD
5, 47–106, 2012