Database of diazotrophs in global ocean
abundances, biomass and nitrogen fixation rates
Luo, Y.-W.; Doney, S.C.; Anderson, L.A.; Benavides, Mar; Berman-Frank, I.; Bode, A.; Bonnet, S.; Boström, K.H.; Böttjer, D.; Capone, D.G.; Carpenter, E.J.; Chen, Y.L.; Church, M.J.; Dore, J.E.; Falcón, L.I.; Fernández, A.; Foster, R.A.; Furuya, K.; Gómez, F.; Gundersen, K.; Hynes, A.M.; Karl, D.M.; Kitajima, S.; Langlois, R.J.; LaRoche, J.; Letelier, R.M.; Maranón, E.; McGillicuddy Jr., D.J.; Moisander, P.; Moore, C.M.; Mourino-Carballido, B.; Mulholland, M.R.; Needoba, J.A.; Orcutt, K.M.; Poulton, A.J.; Rahav, E.; Raimbault, P.; Rees, A.P.; Riemann, Lasse; Shiozaki, T.; Subramaniam, A.; Tyrrell, T.; Turk-Kubo, K.A.; Varela, M.; Villareal, T.A.; Webb, E.A.; White, A.E.; Wu, J.; Zehr, J.P.

Published in:
Earth System Science Data

DOI:
10.5194/essd-4-47-2012

Publication date:
2012

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Luo, Y.-W., Doney, S. C., Anderson, L. A., Benavides, M., Berman-Frank, I., Bode, A., Bonnet, S., Boström, K. H., Böttjer, D., Capone, D. G., Carpenter, E. J., Chen, Y. L., Church, M. J., Dore, J. E., Falcón, L. I., Fernández, A., Foster, R. A., Furuya, K., Gómez, F., ... Zehr, J. P. (2012). Database of diazotrophs in global ocean: abundances, biomass and nitrogen fixation rates. Earth System Science Data, 4, 47-73. https://doi.org/10.5194/essd-4-47-2012
Database of diazotrophs in global ocean: abundance, biomass and nitrogen fixation rates

Y.-W. Luo¹, S. C. Doney¹, L. A. Anderson², M. Benavides³, I. Berman-Frank⁴, A. Bode⁵, S. Bonnet⁶, K. H. Boström⁷, D. Böttjer⁸, D. G. Capone⁹, E. J. Carpenter¹⁰, Y. L. Chen¹¹, M. J. Church¹², J. E. Dore¹², L. I. Falcón¹³, A. Fernández¹⁴, R. A. Foster¹⁵, K. Furuya¹⁶, F. Gómez¹⁷, K. Gundersen¹⁸, A. M. Hynes¹⁹,*, D. M. Karl¹⁰, S. Kitajima¹⁶, R. J. Langlois²⁰, J. LaRoche²⁰, R. M. Letelier²¹, E. Marañón¹⁴, D. J. McGillicuddy Jr.²², P. H. Moisander²²,*, C. M. Moore²³, B. Mouriño-Carballido¹⁴, M. R. Mulholland²⁴, J. A. Needoba²⁵, K. M. Orcutt²⁶, E. Rahav⁴, P. Raimbault⁶, A. P. Rees²⁷, L. Riemann²⁸, T. Shiozaki¹⁶, A. Subramaniam²⁹, T. Tyrrell³⁰, K. A. Turk-Kubo³², M. Varela¹⁵, T. A. Villareal³⁰, E. A. Webb³¹, A. E. White³¹, J. Wu³¹, and J. P. Zehr³²

¹Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA
²Department of Applied Ocean Science and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA
³Instituto de Oceanografía y Cambio Global, Universidad de Las Palmas de Gran Canaria, 35017, Las Palmas de Gran Canaria, Spain
⁴The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan 52900, Israel
⁵Instituto Español de Oceanografía, Centro Oceanográfico de A Coruña, 15080 A Coruña, Spain
⁶IRD-INSU-CNRS, Laboratoire d’Océanographie Physique et Biogéochimique, UMR 6535, Centre d’Océanologie de Marseille, Aix Marseille Université, France
⁷Department of Natural Sciences, Linnaeus University, 39182 Kalmar, Sweden
⁸School of Ocean and Earth Science and Technology, University of Hawaii, Honolulu, Hawaii 96822, USA
⁹Department of Biological Sciences and Wrigley Institute for Environmental Studies, University of Southern California, Los Angeles, California 90089, USA
¹⁰Romberg Tiburon Center, San Francisco State University, Tiburon, California 94920, USA
¹¹Department of Marine Biotechnology and Resources, National Sun Yat-sen University, Kaohsiung 80424, Taiwan
¹²Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, MT 59717, USA
¹³Laboratorio de Ecología Bacteriana, Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico
¹⁴Departamento de Ecoloxía e Bioloxía Animal, Universidade de Vigo, 36310 Vigo, Spain
¹⁵Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, 28359 Bremen, Germany
¹⁶Department of Aquatic Bioscience, University of Tokyo, Yayoi, Bunkyo, Tokyo, 113-8657, Japan
¹⁷Instituto Cavanilles de Biodiversidad y Biología Evolutiva, Universidad de Valencia, P.O. Box 22085, 46071 Valencia, Spain
¹⁸Department of Marine Science, University of Southern Mississippi, Stennis Space Center, Mississippi 39529, USA
¹⁹MIT-WHOI Joint Program in Oceanography/Applied Ocean Science and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA
²⁰Leibniz Institute for Marine Sciences, Duesternbrooker Weg 20, 24105 Kiel, Germany
²¹College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331, USA
²²Ocean Sciences Department, University of California, Santa Cruz, California 95064, USA

Published by Copernicus Publications.
Abstract. Marine N₂ fixing microorganisms, termed diazotrophs, are a key functional group in marine pelagic ecosystems. The biological fixation of dinitrogen (N₂) 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 N₂ fixation rates. Biomass conversion factors are estimated based on cell sizes to convert abundance data to diazotrophic biomass. The database is limited spatially, lacking large regions of the ocean especially in the Indian Ocean. The data are approximately log-normal distributed, and large variances exist in most sub-databases with non-zero values differing 5 to 8 orders of magnitude. Reporting the geometric mean and the range of one geometric standard error below and above the geometric mean, the pelagic N₂ fixation rate in the global ocean is estimated to be 62 (52–73) Tg N yr⁻¹ and the pelagic diazotrophic biomass in the global ocean is estimated to be 2.1 (1.4–3.1) Tg C from cell counts and to 89 (43–150) Tg C from nifH-based abundances. Reporting the arithmetic mean and one standard error instead, these three global estimates are 140 ± 9.2 Tg N yr⁻¹, 18 ± 1.8 Tg C and 590 ± 70 Tg C, respectively. Uncertainties related to biomass conversion factors can change the estimate of geometric mean pelagic diazotrophic biomass in the global ocean by about ± 70%. It was recently established that the most commonly applied method used to measure N₂ fixation has underestimated the true rates. As a result, one can expect that future rate measurements will shift the mean N₂ fixation rate upward and may result in significantly higher estimates for the global N₂ fixation. The evolving database can nevertheless be used to study spatial and temporal distributions and variations of marine N₂ fixation, to validate geochemical estimates and to parameterize and validate biogeochemical models, keeping in mind that future rate measurements may rise in the future. The database is stored in PANGAEA (doi:10.1594/PANGAEA.774851).

1 Introduction

N₂ fixation is the biological conversion of dinitrogen (N₂) gas into two molecules of ammonia by diazotrophic organisms. Over geological time scales, N₂ fixation is important for regulating fixed N concentrations in the ocean and thereby sustaining ocean fertility (Tyrrell, 1999). The rate of pelagic N₂ 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 N₂-fixing microorganisms in the ocean (Karl et al., 2002; Zehr, 2011). However, non-cyanobacterial prokaryotic plankton may also conduct N₂ fixation in the ocean as revealed by the presence and transcription of nifH genes (encoding the iron protein component of the nitrogenase enzyme) (Zehr et al., 1998; Riemann et al., 2010; Farnelid et al., 2011; Fernández et al., 2011), albeit their relative contribution to global N₂ fixation remains to be determined.

There are three major types of marine planktonic diazotrophic cyanobacteria (Sohm et al., 2011b; Zehr, 2011): (1) non-heterocystous filamentous cyanobacteria; (2) heterocystous cyanobacteria; and (3) unicellular cyanobacteria. Among them, most field and laboratory research has focused on non-heterocystous filamentous *Trichodesmium* (*Oscillatoriales*) (Dugdale et al., 1961; Carpenter and Romans, 1991; Capone et al., 1997). *Trichodesmium* is an abundant diazotroph in the open ocean (Capone et al., 2005; LaRoche and Breitbarth, 2005), which can be found as large colonies...
known as “puffs” or “tufts” or as free trichomes (Carpenter, 1983).

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

The most recently characterized diazotrophic phylotypes are the unicellular cyanobacteria (UCYN). Zehr et al. (2001) first successfully amplified a fragment of the nifH gene and nifH transcripts from the <10 µm size fraction of the whole water samples that demonstrated the presence of unicellular diazotrophs. Subsequently, Montoya et al. (2004) measured high rates of N₂ 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., 2012).

Although marine cyanobacterial diazotrophs play a critical role in the oceanic N cycle, primary productivity and organic matter export (e.g., Karl et al., 1997, 2002; Capone, 2000; Gruber, 2008), there is still no database synthesizing the many field measurements of diazotrophic abundances and N₂ 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 N₂ fixation rate of diazotrophs in the global ocean. This effort is part of the Marine Ecosystem Model Intercomparison Project (MAREMIP), in which field measurement-based databases are constructed for biomass and related process rates for phytoplankton functional types (PFTs) (Buitenhuis et al., 2012). The databases, named “Marine Ecosystem DATa” (MAREDAT), include nine PFTs: diatoms, Phaeocystis, coccolithophores, diazotrophs, picophytoplankton, bacterioplankton, mesozooplankton, macrozooplankton and pteropods, and also a database for dissolved organic carbon. In addition to the database for diazotrophs presented in this paper, other databases are presented in other papers of this special volume. The MAREDAT databases will be used for the future model intercomparison studies in the MAREMIP projects, and are also available for public use.

In Sect. 2, the database is described, including information about database construction, data measurement methods, data quality control and conversion from diazotroph abundances to carbon (C) biomass. In Sect. 3, we present and discuss synthesized results from the database, including (1) the results of quality control, (2) spatial and temporal distribution of cell counts of diazotrophs, N₂ fixation rates and nifH-based abundances, (3) general characteristics of the datasets including mean N₂ fixation rates and diazotrophic C biomass as a function of geographical location and depth, (4) estimates for global N₂ fixation rates and diazotrophic C biomass and (5) limitations of the database and the global estimates. In Sect. 4, we draw conclusions and provide recommendations for appropriate use of this database.

2 Data and methods

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

2.1 Database construction

2.1.1 Database summary

Data comprised of three types of direct measurements, cell counts of diazotrophs, N₂ fixation rates, and nifH-based abundances via quantitative polymerase chain reaction (qPCR) assays, were compiled from the scientific literature and personal communication with researchers working in the field. The database contains a total of 12 151 data points including three sub-databases: (1) cell counts of diazotrophs with 5326 data points (Table 1a), (2) N₂ fixation rates with 3624 data points (Table 1b) and (3) diazotrophic abundances estimated from nifH copy abundances (referred to as nifH-based abundances hereafter) with 3201 data points (Table 1c). Note that the counts for Trichodesmium were reported in number of colonies, trichomes or cells, yet we maintain usage of the term “cell counts” to distinguish count-based methods from nifH-based abundances. The diazotroph abundances based on cell counts and nifH genes were converted into C biomass using conversion factors (discussed below). In each sub-database, the data were grouped into three taxonomic types: Trichodesmium, UCYN and heterocystous cyanobacteria (Table 1). A separate grouping is maintained for those N₂ 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., N₂ fixation rates and diazotrophic abundances are considered to be negligible below the bottom depth. These calculated depth-integrated data points were not counted in the total points in Table 1 as
Table 1a. Summary of data points for cell counts of diazotrophs, including volumetric measurements of *Trichodesmium*, unicellular cyanobacteria and heterocystous cyanobacteria and their depth-integrals.

| Investigator | Cell counts for diazotrophs | Region | References |
|--------------|-----------------------------|--------|------------|
|               | *Trichodesmium* | Unicellular | Heterocystous | Depth Integral | Total Points |
| Bar-Zeev     | 26               | 26     | Mediterranean Sea | Bar-Zeev et al. (2008) |
| Benavides    | 29               | 73     | Subtropical N. Atlantic | Benavides et al. (2011) |
| Borstad      | 521              | 521    | Tropical W. Atlantic | Borstad (1978) |
| Brzezinski   | 7                | 7      | Subtropical N. Pacific | Brzezinski et al. (1998) |
| Carpenter    | 377              | 44     | W. Atlantic | Carpenter and Price (1977) |
| Carpenter/Romans | 15         | 15     | Subtropical N. Atlantic | Carpenter and Romans (1991) |
| Capone/Carpenter/Subramaniam | 402           | 176   | Total 101 + 166 = 679 | Capone et al. (2004); Capone et al. (2005) |
| Capone       | 22               | 22     | Arabian Sea | Capone et al. (1998) |
| Chang        | 85               | 146    | East China Sea | Chang et al. (2000) |
| Chen         | 20               | 47     | South China Sea | Chen et al. (2003) |
| Chen         | 20               | 20     | Upstream Kuroshio and South China Sea | Chen et al. (2008) |
| Chen         | 35               | 35     | Northwest Pacific (Kuroshio) | Chen et al. (2011) |
| Dong         | 36               | 36     | South China Sea | Dong et al. (2008) |
| Dugdale      | 15               | 15     | Subtropical N. Atlantic (Subtropical) | Dugdale et al. (1961) |
| Fernández/Mourrío-Carballido/Marañón | 201           | 201   | Cross N.-S. Atlantic | Fernández et al. (2010); Mourrío-Carballido et al. (2011); Fernández et al. (2012) |
| Furuya/Kitajima | 30           | 30     | Tropical/subtropical N. Pacific | Kitajima et al. (2009) |
| Goering      | 51               | 51     | Tropical Atlantic | Goering et al. (1966) |
| Gómez        | 439              | 439    | Tropical/subtropical N. Pacific | Gómez et al. (2005) |
| Gómez        | 83               | 83     | Tropical/subtropical S. Pacific | unpublished data |
| Hegde        | 135              | 135    | Indian Ocean | Hegde et al. (2008) |
| Hulburt      | 19               | 19     | Subtropical N. Atlantic | Hulburt (1962) |
| Hulburt      | 8                | 8      | Subtropical N. Atlantic | Hulburt (1966) |
| Hulburt      | 26               | 26     | Caribbean Sea | Hulburt (1968) |
| Kromkamp     | 44               | 44     | W. Indian Ocean (African coast) | Kromkamp et al. (1997) |
| Lugomela     | 98               | 98     | W. Indian Ocean (African coast) | Lugomela et al. (2002) |
| Letelier/Karl | 23           | 23     | Subtropical N. Pacific (ALOA) | Letelier and Karl (1996) |
| Marumo/Asaoka | 26^b          | 26     | Cross N.-S. Pacific | Marumo and Asaoka (1974) |
| McCarthy/Carpenter | 65       | 13     | Subtropical N. Atlantic | McCarthy and Carpenter (1979) |
| Moore        | 6                | 6      | Tropical/subtropical Atlantic | Moore et al. (2009) |
| Orcutt/Gundersen | 35           | 37     | Subtropical N. Atlantic (BAES) | Orcutt et al. (2001) |
| Poulton      | 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              | 315    | Cross N.-S. Atlantic | Tyrrell et al. (2003) |
| Villareal    | 242              | 242    | Subtropical N. Pacific | Villareal et al. (2011) |
| Villareal    | 186              | 9 + 25 | Subtropical N. Pacific | Villareal et al. (unpublished data) |
| Villareal    | 20               | 30     | Gulf of Mexico | Villareal et al. (unpublished data) |
| White        | 52               | 52     | Gulf of California | White et al. (unpublished data) |
| Wu           | 21               | 3      | South China Sea | Wu et al. (2003) |
| Total        | 3274             | 0      | 1612        | 930            | 5326         |

* Computed from vertical profiles unless marked for those reported by data providers as depth-integrals. ^ Data are reported by data providers as depth-integrated biomass. ^ Do not include depth-integrated profiles unless they are reported by data providers as depth-integrated values (as marked by ^).
Table 1b. Summary of data points for N₂ fixation rates, including volumetric measurements of *Trichodesmium*, unicellular cyanobacteria and heterocystous cyanobacteria and their depth-integrals.

| Investigator | *Trichodesmium* | N₂ Fixation Rates | Region | References |
|--------------|------------------|-------------------|--------|------------|
|              | Unicellular      | Heterocystous     | Whole Seawater | Depth Integrals | Total Points | Method | |
| Bar-Zeev     | 12               | 12                | C₃H₂ reduction | Mediterranean Sea | Bar-Zeev et al. (2008) |
| Benavides    | 53               | 60                | C₃H₂ reduction and C₂N₂ assimilation | Subtropical N. Atlantic | Benavides et al. (2011) |
| Berman-Frank/Rahav | 15      | 1                | 15 C₂N₂ assimilation | Red Sea | Rahav et al. (2012b) |
| Berman-Frank/Rahav | 15      | 6                | 45 C₂N₂ assimilation | Mediterranean Sea | Rahav et al. (2011); Yoge et al. (2011); Rahav et al. (2012a) |
| Bonnet       | 27               | 20                | 47 C₂N₂ assimilation | Tropical Pacific | Bonnet et al. (2009) |
| Bonnet       | 80               | 9                 | 80 C₂N₂ assimilation | Mediterranean Sea | Bonnet et al. (2011) |
| Church/Karl  | 332              | 332               | 332 C₂N₂ assimilation | Subtropical N. Pacific (ALOHA) | Church et al. (2009) |
| Carpenter    | 157              | 39                | 157 C₂H₂ reduction | Tropical/subtropical W. Atlantic | Carpenter and Price (1977) |
| Capone/Carpenter/Subramaniam | 180 | 40 + 33 | 220 C₂H₂ reduction and C₂N₂ assimilation | Subtropical N. Pacific | Sohm et al. (2011a) |
| Capone/Carpenter/Subramaniam | 185 | 7    | 7 C₂N₂ assimilation | Tropical/subtropical N. Atlantic | Capone et al. (2005) |
| Church       | 7                | 7                 | 7 C₂N₂ assimilation | Subtropical N. Pacific | Fong et al. (2008) |
| Dore         | 13               | 24                | 37 C₂N₂ assimilation | Subtropical N. Atlantic | Dore et al. (2002) |
| Falcón       | 24               | 6                 | 24 C₂N₂ assimilation | Subtropical N. Pacific (ALOHA) | Falcón et al. (2004) |
| Fernández/Mourriño-Carballido/Maratón | 115 | 32 | 115 C₂N₂ assimilation | Cross N.-S. Atlantic | Fernández et al. (2010); Mourriño-Carballido et al. (2011); Fernández et al. (2012) |
| Foster       | 1                | 1                 | 1 C₂N₂ assimilation | Subtropical N. Pacific | Needoba et al. (2007) |
| Furuya/Kitajima | 24        | 29               | 53 C₂H₂ reduction | Tropical/subtropical N. Pacific | Kitajima et al. (2009) |
| Furuya/Shiozaki | 103          | 11               | 103 C₂N₂ assimilation | Tropical/subtropical Pacific | Shiozaki et al. (2010) |
| Gandhi       | 28               | 2                 | 28 C₂N₂ assimilation | Indian Ocean | Gandhi et al. (2011) |
| Goering      | 40               | 4                 | 40 C₂N₂ assimilation | Tropical Atlantic | Goering et al. (1966) |
| Gunderson    | 24               | 2                 | 24 C₂H₂ reduction | Subtropical N. Pacific | Gunser et al. (1976) |
| Karl         | 12               | 9                 | 36 C₂N₂ assimilation | Subtropical N. Pacific (ALOHA) | Grabowski et al. (2008) |
| Kromkamp     | 15               | 1                 | 15 C₂H₂ reduction | W. Indian Ocean (African coast) | Kromkamp et al. (1997) |
| Mague        | 8                | 8                 | 8 C₂H₂ reduction | Subtropical N. Pacific | Mague et al. (1974) |
| Mague        | 46               | 46                | 46 C₂H₂ reduction | Subtropical N. Pacific | Mague et al. (1977) |
| McCarthy/Carpenter | 24     | 5               | 24 C₂H₂ reduction | Subtropical N. Atlantic | McCarthy and Carpenter (1979) |
| Montoya      | 16               | 1                 | 16 C₂N₂ assimilation | Subtropical N. Pacific | Montoya et al. (2004) |
| Montoya      | 5                | 5                 | 5 C₂N₂ assimilation | Arafura Sea (near Australia) | Montoya et al. (2004) |
| Moore        | 46               | 10                | 46 C₂N₂ assimilation | Tropical/subtropical Atlantic | Moore et al. (2009) |
| Mulholland   | 30               | 30                | 30 C₂N₂ assimilation | Subtropical NW. Atlantic | Mulholland et al. (2012) |
| Mulholland   | 181              | 181               | 181 C₂N₂ assimilation | Subtropical NW. Atlantic | Mulholland et al. (unpublished data) |
| Mulholland   | 185              | 36                | 185 C₂N₂ assimilation | Subtropical N. Atlantic | Mulholland et al. (unpublished data) |
| Orcutt/Gundersen | 31            | 31               | 31 C₂N₂ assimilation | Subtropical N. Atlantic (BATS) | Orcutt et al. (2001) |
| Raimbault    | 158              | 158               | 158 C₂N₂ assimilation | Tropical/subtropical S. Pacific | Raimbault and Garcia (2008) |
| Raimbault    | 197              | 197               | 197 C₂N₂ assimilation | Subtropical S. Pacific | Garcia et al. (2007) |
| Raimbault    | 71               | 71                | 71 C₂N₂ assimilation | Mediterranean Sea | Sandroni et al. (2007) |
| Rees         | 5                | 5                 | 5 C₂N₂ assimilation | Eastern Mediterranean Sea | Rees et al. (2006) |
| Rees         | 21               | 21                | 21 C₂N₂ assimilation | Subtropical N. Atlantic (FeeP) | unpublished data |
| Rees         | 70               | 70                | 70 C₂N₂ assimilation | Cross N.-S. Atlantic (AMTI4&15) | unpublished data |
| Rees         | 2                | 2                 | 2 C₂N₂ assimilation | Western English Channel | Rees et al. (2009) |
| Rees         | 57               | 5                 | 57 C₂N₂ assimilation | Tropical E. Atlantic | Turk et al. (2011) |
| Turk/Zehr    | 11               | 11                | 11 C₂N₂ assimilation | Tropical Atlantic | Goebel et al. (2010) |
| White/Letelier | 11            | 12               | 23 C₂N₂ assimilation | Subtropical N. Pacific | White et al. (unpublished data) |
| White/Letelier | 10            | 43               | 53 C₂N₂ assimilation | Subtropical N. Pacific | Watkins-Brandt et al. (2011) |
| White        | 134              | 19                | 134 C₂N₂ assimilation | Gulf of CA and Tropical Pacific | White et al. (2011) |

Data are reported by data providers as depth-integrated N₂ fixation rates.

**Notes:**
- Do not include integrated profiles unless they are reported by data providers as depth-integrated N₂ fixation rates.
- Computed from vertical profiles unless marked by an asterisk (*) unless they are reported by data providers as depth-integrated values (as marked by °).
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 | Heterocystous | Depth Integral | Total Points<sup>a</sup>   |
| Bonnet                        | 12 23 12 47          | Tropical/subtropical S. Pacific | Bonnet et al. (2008) |
| Boström/Riemann Church        | 5 16 58 58           | Baltic Sea                    | Boström et al. (2007) |
| Church                        | 24 48 24 8 40        | Subtropical N. Pacific (ALOHA) | Church et al. (2005a) |
| Church                        | 8 16 8 50<sup>b</sup> | Subtropical N. Pacific (ALOHA) | Fong et al. (2005b)  |
| Foster/Zehr                   | 66 198 66 300        | Subtropical Atlantic          | Foster et al. (2007) |
| Foster/Needoba/Zehr           | 18 18 18             | Subtropical N. Pacific        | Needoba et al. (2007) |
| Foster/Zehr                   | 6 18 18              | Red Sea                       | Foster et al. (2009a) |
| Foster/Subramaniam/Zehr       | 19 40 18 18          | Tropical E. Atlantic          | Foster et al. (2009b) |
| Foster/Zehr                   | 17 42 17 14          | Gulf of California            | White et al. (unpublished data) |
| Liu Langlois/LaRoche          | 25 49 14 15          | South China Sea               | Kong et al. (2011)  |
| Mazarid/Scanlan/Moisander/Zehr| 54 105 36 18         | Arabian Sea                   | Lazaro et al. (2008) |
| Moisander/Zehr                | 178 367 110 110      | South China Sea               | Moisander et al. (2004) |
| Mulholland                    | 22 51 37 110         | Subtropical Atlantic          | Moisander et al. (2010) |
| Orcutt/Gundersen              | 18 18 18             | Gulf of Mexico                | Mulholland et al. (2012) |
| Rees/Turk                     | 29 58 27 114         | Tropical E. Atlantic          | Orcutt et al. (unpublished data) |
| Turk/Zehr                     | 154 300 154 154      | Tropical Atlantic             | Turk et al. (2011)  |
| Total                         | 759 1793 599 416     | 3201                          |                     |

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

they are derived values. An additional 709 data points of diazotroph cell counts and N₂ fixation rates that were originally reported as depth-integrated values were also included. These 709 depth-integrated data points are counted in the total in Table 1 as they are independent from other data points. Each depth-integrated data point is identified by its sampling date, geographic location and integral depth.

The database also provides total diazotrophic C biomass (from cell counts and *nifH*-based abundances) for each sample by summing values from the three defined diazotrophic types: *Trichodesmium*, unicellular and heterocystous cyanobacteria. Total N₂ fixation rates are also provided: when whole seawater N₂ fixation rates are available, they are used as total N₂ fixation rates. Otherwise, the total N₂ fixation rates are calculated by summarizing values from the three defined diazotrophic types. In many samples, measurements were not available for all the three defined diazotrophic types. Also, these three defined types do not represent the full diazotrophic community. Thus, the derived totals (via summation) can be considered as the lower limits of diazotrophic biomass and activity.

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

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

Most counts for Trichodesmium were reported in number of colonies or trichomes per volume, and in a few datasets in cell densities. In order to use a unified biomass conversion factor for Trichodesmium (discussed later), all the Trichodesmium counts were converted to number of trichomes assuming commonly used conversion factors of 200 trichomes colony⁻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 (Bruzielinski et al., 1998; Gómez et al., 2005; Poulton et al., 2009; Villareal et al., 2011; Villareal et al., unpublished Gulf of Mexico data) in which abundances of host diatom Hemiaulus and Rhizosolenia were reported while Richelia heterocysts were not counted. As a Hemiaulus diatom typically contains 2 Richelia filaments whereas a Rhizosolenia diatom can contain 1–32 Richelia filaments (Sundström, 1984; Villareal, 1989, 1990; Foster and O’Mullan, 2008), the abundances of heterocysts for these datasets were derived from cell counts of their host diatoms by assuming that each Hemiaulus or Rhizosolenia cell contains 2 or 5 Richelia filaments, respectively. In one dataset (Gómez et al., 2005), abundances of Chaetoceros were counted but the associated heterocystous cyanobacteria were found to be Richelia. An average ratio of 0.5 Richelia heterocyst per Chaetoceros cell was reported by the dataset and was used to calculate the Richelia abundance. The trichomes or filaments of Richelia or Calothrix are typically composed of 3–4 vegetative cells and 1 terminal heterocyst (Foster and Zehr, 2006). Thus, abundances of cells within these genera are estimated by multiplying the heterocyst abundances by 5, i.e., assuming 5 cells per filament. Note that it may underestimate heterocystous cell abundances as the Richelia symbiont of Rhizosolenia in some cases can contain more vegetative cells (near 10) (Villareal, 1989, 1992; Janson et al., 1999).

2.1.3 N₂ fixation rates

N₂ fixation rates were measured directly by ¹⁵N-labeled N₂ gas (¹⁵N₂) assimilation (Montoya et al., 1996) or indirectly by the acetylene (C₂H₂) reduction assay (Capone, 1993) (Table 1b). The ¹⁵N₂ assimilation method tracks the conversion of ¹⁵N₂ to particulate N. The ¹⁵N₂ tracer is added into the ambient pool of N₂; the ¹⁵N/¹⁴N ratio is measured in the particulate N after incubation and compared to the natural abundance of N isotopes in unlabeled particulate material. The C₂H₂ method estimates N₂ fixation rate indirectly by measuring the reduction of C₂H₂ (a competitive inhibitor of N₂) to ethylene (C₂H₄) which is then converted to a N₂ fixation rate assuming 3 or 4 moles C₂H₂ reduced per 1 mole N₂ fixed, depending on the extent of nitrogenase-linked hydrogen production (Postgate, 1998). Generally, the direct ¹⁵N₂ assimilation is a precise and sensitive method; hence, it has been used to generate the majority of rates (Table 1b). Direct comparison of these two methods showed the direct ¹⁵N₂ assimilation method generally yields lower rates than those estimated from C₂H₂ reduction assay (see summary in Mulholland, 2007). Discrepancies between these measures could be because the ¹⁵N₂ assimilation method measures the net rate of conversion of reduced N to cellular N, or net N₂ fixation, while the C₂H₂ reduction method measures gross N₂ 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 ¹⁵N₂ assimilation method significantly underestimates the N₂ fixation rates because the ¹⁵N₂ bubbles injected in seawater do not attain equilibrium with surrounding water (Mohr et al., 2010), which will be discussed later. We include the N₂ fixation rates acquired by either the ¹⁵N₂ assimilation method or the C₂H₂ reduction assays in the database, considering that they distribute within a similar range of magnitude (Fig. 1). But this analysis does not consider different sampling sites. Further investigations using pair-wise comparison of the methods are needed to evaluate effects of these two methods on N₂ fixation rate measurement. Users have to be careful when using the database to study N₂ fixation rates aggregated from two different methods.

The collected N₂ fixation rates were mostly measured for whole seawater samples (Table 1b). Some samples were filtered and N₂ fixation rates were measured for organisms in the <10µm 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 \(<10\mu m\) fractions. N\(_2\) fixation rates were also measured in some datasets specifically for *Trichodesmium* and heterocystous cyanobacteria. Most N\(_2\) fixation data were reported as hourly rates, except for 11 datasets that were reported as daily rates. *Trichodesmium* fixes N\(_2\) exclusively during the light period, while the diel patterns of N\(_2\) fixation are unclear for other diazotrophs (Carpenter and Capone, 2008 and references therein). Thus, those hourly N\(_2\) fixation rates were converted to daily rates by multiplying by 12 h, which, however, could be conservative if diazotrophs other than *Trichodesmium* could fix N\(_2\) during night (e.g., Montoya et al., 2004; Zehr et al., 2007).

### 2.1.4 NifH-based abundances

NifH abundances were estimated by qPCR targeting the *nifH* gene (Church et al., 2005a, b; Foster et al., 2007). Cellular DNA was extracted, and the gene sequences were targeted for different diazotrophic groups. Most *nifH*-based abundances were estimated for the three major diazotroph types: *Trichodesmium*, UCYN groups and heterocystous groups (Table 1c). Gene copies for UCYN were identified as UCYN-A, UCYN-B and/or UCYN-C groups. *NifH*-based abundances were also estimated for different groups of heterocystous cyanobacteria based on three *nifH* gene sequences (het-1, het-2 and het-3), which have been identified in symbioses with diatoms: *Richelia-Rhizosolenia, Richelia-Hemiaulus* and *Calothrix-Chaetoceros*, respectively (Church et al., 2005b; Foster and Zehr, 2006). Note that there is one dataset (Boström et al., 2007) that reports the abundance of heterocystous genus *Nodularia*. The sensitivity of the qPCR thermocycler is usually in the range of 6–12 *nifH* copies per reaction. When accounting for sample filtration volume, final elution volume and the volume of the DNA extract used in the qPCR reaction, the detection limit of the assay is significantly higher. Due to the variety of factors that determine the assay detection limit, it varies between laboratories and even datasets and is typically on the order of 25–100 *nifH* copies l\(^{-1}\). Thus, for those data points reported as “detected but not quantifiable”, 10 *nifH* copies l\(^{-1}\) could be a conservative estimate.

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

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

### 2.2 Log-normal distribution and quality control

Diazotrophic abundances and N\(_2\) fixation rates in the ocean can range from 0 when diazotrophs are below detection or truly absent at that location and time, to very high values during diazotrophic bloom phases. Cell abundances and N\(_2\) fixation rates hence vary by several orders of magnitude, are often not normally distributed and are positively skewed (long tail of high values). However, the datasets (excluding...
Table 2. Estimated default and the upper and lower bounds for the biomass conversion factors, and their impacts on global biomass estimates (based on geometric mean).

|                          | Conversion Factor (pg C cell\(^{-1}\)) | Global Biomass Estimate\(^a\) (Tg C) |
|--------------------------|----------------------------------------|-----------------------------------|
|                          | default lower upper                    | default lower upper                |
| Cell-count-based biomass  |                                        |                                   |
| Trichodesmium\(^b\)      | 300 100 500                            |                                   |
| Richelia                 | 10 2 80                                |                                   |
| Calothrix                | 10 5 20                                |                                   |
|                          |                                        | 2.1 0.8 3.5                        |
| 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          |                                        |                                   |
| Richelia                 | 10 2 80                                |                                   |
| Calothrix                | 10 5 20                                |                                   |
|                          |                                        | 89 28 170                           |

\(^a\) The low/high biomass is calculated when all their subtypes use low/high conversion factors.

\(^b\) Assuming 100 cells trichomes\(^{-1}\).

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

We control for data quality by using Chauvenet’s criterion to flag suspicious outliers (Glover et al., 2011), which generally applies to normally distributed datasets and rejects data whose probability of deviation from the mean is less than 1/(2\(n\)), where \(n\) is the number of data points. However, considering (1) the datasets are log-normally distributed and (2) valid diazotrophic abundances and N\(_2\) fixation rates can be infinite low or zero, Chauvenet’s criterion in our practice is applied to the log-transformed non-zero data to flag the outliers only on the high side. Those \(nifH\)-based abundances that were reported as “detected but not quantifiable” are not included in the application of Chauvenet’s criterion. Although not used in the application of Chauvenet’s criterion, data with zero values are kept in the database, as they represent valuable ecological information. Note that the criterion is processed separately for the volumetric data points of each taxonomic type. The criterion is also applied to the depth-integrated total N\(_2\) fixation rates and the depth-integrated total biomass estimated from cell counts and \(nifH\)-based abundances, as these depth-integrated values are used later for the global estimates. First, the mean \(\bar{x}_\text{log}\) and the standard deviation \(\sigma_{\text{log}}\) of the log-transformed data are calculated, which are used to calculate the critical value \(x_{\text{log}}^*\) with a probability of a half of 1/(2\(n\)) that values would exceed the mean \(\bar{x}_\text{log}\) by this amount assuming normal distribution (in log-transformed space). One-half of 1/(2\(n\)) 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 \(\bar{x}_\text{log} + x_{\text{log}}^*\) are flagged.

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

2.3 Biomass conversion

The cell counts and \(nifH\)-based abundances are converted to C biomass using conversion factors (Table 2). As discussed above, for C biomass estimates all the counts of \(Trichodesmium\) in colonies and cells have been converted to number of trichomes by assuming 200 trichomes colony\(^{-1}\).
Figure 2. Histogram of data points on logarithmic scale (blue bars) and the critical values for quality control using Chauvenet’s criterion (dashed red lines). Values higher than the critical values are rejected. (a) *Trichodesmium* cell counts, (b) *Trichodesmium* $N_2$ 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.
and 100 cells trichome$^{-1}$, which were widely used as an estimate for *Trichodesmium* (LaRoche and Breitbarth, 2005). To determine the biomass conversion factor for trichomes of *Trichodesmium*, we utilize size measurements of cultured cells of different *Trichodesmium* species (Hynes et al., 2012) and estimate carbon content using the model of Verity et al. (1992) (Tables 3 and S1). The format of the Verity et al. (1992) model is

\[ C = 0.433 \cdot V^{0.863}, \]

where \( C \) is the cell carbon content in pg and \( V \) is the cell volume in \( \mu \text{m}^3 \). This model was used because it was based on many data points across a wide range of cell sizes and species type, including cyanobacteria. *T. erythraeum* is the smallest *Trichodesmium* species, with estimated \( C \) of 65 pg C cell$^{-1}$ (Table 3) or 6.5 ng C trichome$^{-1}$ (using 100 cells trichome$^{-1}$), which is consistent with the back-calculated 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). Thus, the Verity et al. (1992) model used here appears to be suitable for *Trichodesmium*. However, *T. erythraeum* is relatively small compared to other *Trichodesmium* species. The estimated carbon content for other *Trichodesmium* species is 110–250 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 \( \mu \text{g} \) C colony$^{-1}$ in the Pacific (Mague et al., 1977) and 10.9 and 11.6 \( \mu \text{g} \) C colony$^{-1}$ for *Trichodesmium* puffs and tufts in Atlantic (McCarthy and Carpenter, 1979), which results in a conversion factor of 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 \( \mu \text{g} \) C colony$^{-1}$ \((n = 23)\) for *Trichodesmium* puffs, 9.9 ± 2.5 \( \mu \text{g} \) C colony$^{-1}$ \((n = 14)\) for *Trichodesmium* tufts and 4.5 ± 2.4 \( \mu \text{g} \) C colony$^{-1}$ \((n = 6)\) for *T. erythraeum* (Orcutt and Gundersen, unpublished data), which, by assuming 200 trichomes colony$^{-1}$, leads to a conversion factor of 41, 49 and 22 ng C trichome$^{-1}$, respectively. The synthesis of laboratory and field data by LaRoche and Breitbarth (2005) provides values of 50–58 ng C trichome$^{-1}$.

To accommodate all these estimates, we use default conversion factors of 300 pg C cell$^{-1}$ (30 ng C trichome$^{-1}$), using 100 cells trichome$^{-1}$ for *Trichodesmium*, with its range estimated to 100–500 pg C cell$^{-1}$ (Table 2).

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

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

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

| Species         | Biovolume (\( \mu \text{m}^3 \)) | Carbon content (pg C cell$^{-1}$) |
|-----------------|----------------------------------|----------------------------------|
|                 | mean ± standard deviation | range                           | mean ± standard deviation | range                           |
| *T. thiebautii*  | 680 ± 380                       | 280–1200                         | 120 ± 57                    | 55–190                          |
| *T. tenue*       | 610                             |                                   | 110                         |                                 |
| *T. pelagica*    | 1100 ± 130                      | 1000–1200                        | 190 ± 18                    | 170–200                        |
| *T. hildebrandii*| 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                        |

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.
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 Biovolume (μm³) | Heterocyst C content (pg C cell⁻¹) | Vegetative cell Biovolume (μm³) | Vegetative cell C content (pg C cell⁻¹) | Vegetative cells per trichome | Biomass conversion (pg C cell⁻¹) |
|-----------|-----------------------------|----------------------------------|-------------------------------|----------------------------------------|------------------------------|---------------------------------|
|           |                             |                                  |                               |                                        |                              |                                 |
| Richelia  | 110 ± 180                   | 22 ± 31                          | 42 ± 76                       | 9 ± 14                                 | 3                            | 13 ± 18                         |
|           |                             |                                  |                               |                                        | 5                            | 12 ± 16                         |
|           |                             |                                  |                               |                                        | 10                           | 11 ± 15                         |
| Calothrix | 45 ± 35                     | 11 ± 8                           | 47 ± 19                       | 12 ± 4                                 | 3                            | 12 ± 5                          |
|           |                             |                                  |                               |                                        | 5                            | 12 ± 5                          |
|           |                             |                                  |                               |                                        | 10                           | 12 ± 4                          |

equals 90–160 μm³ and leads to a conversion factor of 15–24 pg C cell⁻¹ by using the Verity et al. (1992) model. By merging these two estimates, the default conversion factor of 10 pg C cell⁻¹ with a range of 5–24 pg C cell⁻¹ is used for UCYN-C (Table 2).

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

3 Results and discussion

3.1 Results of quality control

Most data types are well approximated by a log-normal distribution (Figs. 2 and S1), except for Calothrix cell counts and Calothrix nifH-based abundances which only have limited non-zero data points (Fig. S1b and h). By applying Chauvenet’s criterion, there are only 9 data points flagged, including 1 volumetric Trichodesmium cell count (Fig. 2a), 1 volumetric Trichodesmium N₂ fixation (Fig. 2b), 2 volumetric N₂ fixation rates by UCYN (Fig. 2c), 2 volumetric whole seawater N₂ fixation rate (Fig. 2d) and 3 volumetric UCYN-B nifH genes (Fig. 2e). Thus from a statistical point of view, most data are acceptable. The flagged 1 data point of Trichodesmium cell count of 4.4 × 10¹⁰ trichomes m⁻³ and 1 data point of Trichodesmium N₂ 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 N₂ fixation rate were observed, these two data points were very likely real. The flagged data points of UCYN’s N₂ fixation rate of 360 and 960 μmol N m⁻³ d⁻¹ (Montoya et al., 2004) were from ~10° S, ~130–135° E, the western tropical Pacific in the Arafura Sea (Fig. 3b). High N₂ fixation rates are commonly found in this region, such as another flagged whole seawater N₂ 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 N₂ fixation rates (Bonnet, unpublished data). The possible reason for these high N₂ fixation rates could be the iron supply from volcanoes or via upwelling. Another flagged whole seawater N₂ fixation rate of 13,500 μmol N m⁻³ d⁻¹ (Gandhi et al., 2011) was sampled from surface water at ~17° N, 73° E, near the Indian coast (Fig. 3b). In the same vertical profile, the N₂ fixation rates sampled in 2 m and deeper were all less than 32 μmol N m⁻³ d⁻¹, and all the other surface N₂ fixation rates sampled in the surrounding area on this cruise were below 540 μmol N m⁻³ d⁻¹. We evaluate that this high rate is extraordinary and should be removed from the database. The flagged three data points of UCYN nifH-based abundances of ~2–9 × 10¹¹ 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 three flagged data points of the extreme peaks in nifH-based abundance were sampled in June and July 2009 during extremely high water temperature events (~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 increases towards 3–6 by the end of the late summer period (likely due to N\textsubscript{2} fixation). The extreme peaks in nifH-based abundance also coincided with peaks in phytoplankton chlorophyll a (5–6 µg l\textsuperscript{-1}). Thus, we believe most flagged data points, except the high surface N\textsubscript{2} fixation rate found near the Indian coast (Gandhi et al., 2011), are due to the specific environment and not necessarily related to the data quality. Hence, we have retained these values in the database. These points, however, are not included in our later analyses because their extremely high values would influence the mean values. It is also important to note that we have not excluded data based on an
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 for four datasets in the Arabian Sea (Capone et al., 1998; Mazard et al., 2004), near the Kenyan coast (Kromkamp et al., 1997) and in the Madagascar Basin (Poulton et al., 2009). There are also some data points in inner seas. N$_2$ fixation rates were measured almost every month at the BATS station in 1995–1997 and at Station ALOHA from 2005–2010. These are the only two sites with long-term sustained time series of N$_2$ fixation measurements.

Most data were collected in the tropical and subtropical regions, with latitudinal coverage of 50°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 March, April and July (Fig. 5f).

3.3 N$_2$ fixation rates

After being binned onto 3° × 3° 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 µmol N m$^{-2}$ d$^{-1}$ (Fig. 6a). In most other regions, depth-integrated N$_2$ fixation rates are on an order of 1–100 µmol N m$^{-2}$ d$^{-1}$ (Fig. 6a).

The volumetric N$_2$ fixation rates are also analyzed on 3° × 3° grids in the five vertical layers of 0–5 m, 5–25 m, 25–62.5 m, 62.5–137.5 m and 137.5–250 m. The geometric mean values of each grid box are illustrated in Fig. 6b–f. Note that the depth-integrated and volumetric values are not from exactly the same data sources. They overlap for some data sources but also have locations that are not in common, as some of the data were originally reported as depth-integrated values and the other depth-integrated values were calculated herein. N$_2$ fixation rates generally decrease with depth (Fig. 6b–f). When compared horizontally in each layer, the database reveals that there are several regions with high N$_2$ fixation rates: the subtropical North Pacific in all the layers, surface waters of the western Pacific and the tropical Atlantic (especially in the west) in 0–25 m (Fig. 6b–f).

3.4 Diazotrophic abundance and biomass

Cell counts and nifH-based abundances are used to estimate diazotrophic C biomass using the default biomass conversion factors (Table 2), except for a few datasets where the
Figure 5. Temporal distribution of the data points in year and month for (a)–(b) cell counts for diazotrophs, (c)–(d) N₂ fixation rates, and (e)–(f) nifH-based abundances.

contributors measured the conversion factors. The cell count data demonstrate that *Trichodesmium* is the dominant diazotroph, except that the abundance of *Calothrix* can be as large as *Trichodesmium* (Table 5). But the number of depth-integrated cell count samples for *Calothrix* are very limited (Table 5) and are limited to one cruise in the subtropical North Pacific (Villareal et al., unpublished). As the average *Trichodesmium* cell size, and thus the biomass 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 shows that *nifH*-based abundances are mostly one order of magnitude higher than cell-count-based abundances for *Trichodesmium* and heterocystous cyanobacteria, except that the volumetric *Trichodesmium* abundances are comparable in both cell-count-based and *nifH*-based geometric means (Table 5). In the North Atlantic Ocean where both cell counts and *nifH*-based abundances were frequently measured (Fig. 3a and c), histograms of *Trichodesmium* abundances derived from both the cell count data (assuming 100 cells trichome⁻¹) and *nifH*-based data are in agreement (Fig. 7). This evidence provides some support for the reliability of *nifH*-based abundance for *Trichodesmium*.

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

Total diazotrophic C biomass estimated from cell counts is displayed in spatial maps for a given depth-integral and for the five vertical layers of 0–5 m, 5–25 m, 25–62.5 m, 62.5–137.5 m and 137.5–250 m by showing the geometric means of each 3° × 3° grid bin (Fig. 8a–f). The depth-integrated cell-count-based diazotrophic biomass is higher in the western than in the eastern Atlantic (Fig. 8a). The depth-integrated cell-count-based diazotrophic biomass is low in all layers (Fig. 8b–f). The high
Figure 6. (a) Geometric mean depth-integrated N\textsubscript{2} fixation rates binned on 3\textdegree\times3\textdegree grid. Also geometric mean N\textsubscript{2} fixation rates on 3\textdegree\times3\textdegree 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.

3.5 First-order estimates for global N\textsubscript{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\textsubscript{2} fixation rate and diazotrophic biomass. We divide the global ocean into 6 regions: the North and the South Atlantic Ocean, the North and the South Pacific Ocean, the Indian Ocean and the Mediterranean Sea (Tables 6, 7 and 8). Two methods, geometric mean and arithmetic mean, were used for the estimation. As the data are not evenly distributed in space and intensive samplings were made in some regions (Fig. 3), the depth-integrated values are first binned to 3\textdegree\times3\textdegree 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\textsubscript{2} fixation...
| Depth-integrated Data | n | n of zero-value data | Abundance (10^9 cells m^{-2}) | Biomass (mg C m^{-2}) |
|-----------------------|---|----------------------|-------------------------------|----------------------|
|                       |   |                      | Geometric                     | Arithmetic           |
|                       |   |                      |                               |                      |
| cell-count-based data  |   |                      |                               |                      |
| Trichodesmium         | 620| 288                 | 0.16 (0.14–0.18)              | 1.2 (0.12)           |
|                       |   |                      |                               |                      |
| Richelia              | 288| 17                  | 0.010 (0.008–0.013)           | 3.3 (0.88)           |
| Calothrix             | 17 | 9                   | 0.084 (0.050–0.14)            | 0.096 (0.050)        |
|                       |   |                      |                               |                      |
| nifH-based data       |   |                      |                               |                      |
| Trichodesmium         | 103| 10                  | 1.4 (1.1–1.9)                 | 8.8 (1.5)            |
| UCYN-A                | 104| 24                  | 1.5 (0.88–2.4)                | 3400 (2200)          |
| UCYN-B                | 103| 20                  | 0.11 (0.071–0.16)             | 3.1 (0.93)           |
| UCYN-C                | 31 | 18                  | 0.055 (0.020–0.16)            | 0.32 (0.16)          |
| Richelia              | 63 | 2                   | 0.24 (0.15–0.38)              | 0.096 (0.050)        |
| Calothrix             | 11 | 0                   | 0.21 (0.093–0.46)             | 0.84 (0.50–1.4)      |

Volumetric Data | n | n of zero-value data | Abundance (10^9 cells m^{-3}) | Biomass (µg C m^{-3}) |
|----------------|---|----------------------|-------------------------------|----------------------|
|                |   |                      | Geometric                     | Arithmetic           |
|                |   |                      |                               |                      |
| cell-count-based data |   |                      |                               |                      |
| Trichodesmium  | 3222| 1094               | 3.3 (3.1–3.5)                 | 29 (3.6)             |
| Richelia       | 1564| 781                | 0.11 (0.097–0.12)             | 11 (1.5)             |
| Calothrix      | 48 | 28                 | 0.091 (0.060–0.14)            | 0.13 (0.043)         |
|                |   |                      |                               |                      |
| nifH-based data |   |                      |                               |                      |
| Trichodesmium  | 759| 163                | 2.8 (2.5–3.2)                 | 87 (12)              |
| UCYN-A         | 771| 334                | 5.5 (4.5–6.8)                 | 23 000 (9,400)       |
| UCYN-B         | 792| 322                | 1.1 (0.95–1.3)                | 280 (75)             |
| UCYN-C         | 226| 141                | 1.2 (0.90–1.7)                | 4.8 (1.1)            |
| Richelia       | 535| 160                | 1.6 (1.3–1.9)                 | 240 (63)             |
| Calothrix      | 64 | 52                 | 0.33 (0.17–0.62)              | 0.24 (0.10)          |

* Assuming 100 cells trichome^{-1}.

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, 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\textsubscript{2} fixation rates were also low in the South Atlantic (Table 6). There are inadequate depth-integrated N\textsubscript{2} fixation rates collected in the Indian Ocean to estimate the means of this basin. By summing the geometric mean rates from all the regions, the global N\textsubscript{2} fixation rate (not including the Indian Ocean) is estimated to be 62 (error range: 53–73) Tg N yr\textsuperscript{-1} (Table 6), which is at the low end of the current geochemical estimates of 100–200 Tg N yr\textsuperscript{-1} for marine pelagic N\textsubscript{2} fixation (Gruber and Sarmiento, 1997, 2002; Karl et al., 2002; Galloway et al., 2004; Deutsch et al., 2007; Gruber, 2008). Note that, as described in Sect. 2.2, the error range for a geometric mean in this study is represented as between the geometric mean multiplied/divided by the corresponding geometric standard error. The estimate of global N\textsubscript{2} fixation rate using arithmetic mean is higher at 140 (standard error: 9.2) Tg N yr\textsuperscript{-1} (Table 6), which is within the range of the current geochemical estimates.

In a manner similar to our global estimates of the N\textsubscript{2} fixation rate, we have also calculated global mean diazotrophic biomass from cell-count-based biomass data (Table 7). As in many sampling sites in the Pacific Ocean, only heterocystous cyanobacteria are counted (Table 1a) and thus very low total diazotrophic biomass was reported for these sites (Fig. 8a). We give special consideration for the Pacific by estimating mean biomass for Trichodesmium and heterocystous cyanobacteria separately before summing the values from these two taxonomies (Table 7). The average diazotrophic biomass in the North Atlantic is similar to or higher than that in the North Pacific as shown by both geometric and arithmetic means (Table 7). The global diazotrophic biomass (not including Mediterranean Sea) is estimated from the cell-count-based data as 2.1 (error range: 1.4–3.1) Tg C using geometric mean or 18 (standard error: 1.8) Tg C using arithmetic mean (Table 7).

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

By comparing the global estimate of N\textsubscript{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\textsubscript{2} fixation is estimated to be 2 days and 92 days, respectively.

3.6 Limitations

There are a number of limitations with the database and the global estimates. First and foremost, these data points are not uniformly distributed in the world ocean. Historically, scientists have also sought out regions with high diazotrophic abundances and, hence, there is a higher possibility of artificially elevated N\textsubscript{2} fixation activity and diazotrophic biomass relative to true regional means. As the current database does not cover some regions such as the coastal upwelling zones,
our estimate could be changed substantially if values differ within these biogeographic domains.

As mentioned above, it was recently established that the most commonly applied method used to measure N$_2$ fixation, the $^{15}$N$_2$ assimilation method (Table 1b), has underestimated the true rates because the $^{15}$N$_2$ bubbles injected in seawater do not attain equilibrium with surrounding water (Mohr et al., 2010). Two independent method comparisons conducted in the Pacific and Atlantic Oceans demonstrated that the measured rates are significantly higher when the $^{15}$N$_2$ tracer is added in a dissolved form to the incubations, resulting in an instantaneous $^{15}$N isotopic equilibrium in the seawater (Großkopf et al., 2012; Wilson et al., 2012). The variability in the underestimation was linked to the composition of the diazotrophic community (Mohr et al., 2010; Großkopf et al., 2012) indicating that it will be difficult to correct the historical rate measurements made with the $^{15}$N$_2$ method. However, these measurements can be treated as minimum estimates that may increase in the future. These measurements are still valuable when comparing N$_2$ fixation rates in different regions, as they are approximately log-normally distributed and range in ~6 orders of magnitude (Fig. 1).

It is also a notable issue that the difference between the geometric and the arithmetic mean N$_2$ fixation rate is larger in the North Atlantic than those in the other basins (Table 6). A strict log-normal distribution is symmetric in the log space and its geometric mean is located at the peak of the distribution, i.e., at the highest probability. However, the depth-integrated total N$_2$ fixation rates in our database, after binned on $3^\circ \times 3^\circ$ grid, are not strictly log-normally distributed in each basin (Fig. 10). The North Atlantic has highest spatial coverage, and the N$_2$ fixation rates in this basin differ in 6 orders of magnitude from $10^{-3}$–$10^{3}$ µmol N m$^{-2}$ d$^{-1}$ (Fig. 10). Its distribution is not symmetric, with more values observed on the left (lower) side. Thus, the geometric mean N$_2$ fixation rate for the North Atlantic is one order of magnitude lower.

Figure 8. (a) Geometric mean depth-integrated cell-count-based diazotrophic carbon biomass binned on $3^\circ \times 3^\circ$ grid. Also geometric mean cell-count-based diazotrophic carbon biomass on $3^\circ \times 3^\circ$ grid for 5 vertical layers of (b) 0–5 m, (c) 5–25 m, (d) 25–62.5 m, (e) 62.5–137.5 m and (f) 137.5–250 m. The color bar is in logarithmic scale. Zero values are marked with black triangles.
Figure 9. (a) Geometric mean depth-integrated $nifH$-based diazotrophic carbon biomass binned on $3^\circ \times 3^\circ$ grid. Also geometric mean $nifH$-based diazotrophic carbon biomass on $3^\circ \times 3^\circ$ grid for 5 vertical layers of (b) 0–5 m, (c) 5–25 m, (d) 25–62.5 m, (e) 62.5–137.5 m and (f) 137.5–250 m. The color bar is in logarithmic scale. Zero values are marked with black triangles.

Table 6. Estimates of N$_2$ fixation rate for the global oceans based on geometric and arithmetic means. Note that data points with zero-values are excluded when calculating geometric means. Depth-integrated N$_2$ fixation rates are first binned to $3^\circ \times 3^\circ$ grid, and geometric and arithmetic means are calculated for each bin, which are then used to calculate geometric and arithmetic means for each region. Areal sum is calculated by multiplying geometric or arithmetic means with ocean area. Error ranges for geometric means shown in parentheses are estimated by dividing and multiplying the geometric means by one geometric standard error. Standard errors for arithmetic means are shown in parentheses. Error range for global geometric mean is estimated by simply summing lower and upper bounds of the error ranges of each region. Errors for global arithmetic means are propagated from (independent) errors in each region. ND: no data. NQ: not qualified to calculate areal sum because of too few samples.

| Region                  | Latitudinal range | Ocean area ($\times 10^{12}$ m$^2$) | $n$ of bins | $n$ of zero-value bins | N$_2$ fixation rate (umol N m$^{-2}$ d$^{-1}$) Geometric | Areal Sum (Tg N yr$^{-1}$) Geometric | Areal Sum (Tg N yr$^{-1}$) Arithmetic |
|-------------------------|-------------------|-------------------------------------|-------------|------------------------|-----------------------------------------------------|-------------------------------------|-------------------------------------|
| North Atlantic Ocean    | 0–55° N           | 36                                  | 125         | 0                      | 9.2 (7.0–12)                                        | 170 (40)                            | 1.7 (1.3–2.2)                      |
| South Atlantic Ocean    | 40–0° S           | 27                                  | 15          | 0                      | 7.9 (6.3–9.8)                                       | 13 (4.4)                            | 1.1 (0.9–1.4)                      |
| North Pacific Ocean     | 0–55° N           | 89                                  | 45          | 3                      | 78 (67–90)                                          | 120 (22)                            | 35 (30–41)                         |
| South Pacific Ocean     | 40–0° S           | 72                                  | 26          | 0                      | 64 (54–76)                                          | 130 (46)                            | 24 (20–28)                         |
| Indian Ocean            | 40° S–25° N       | 62                                  | 4           | 0                      | 120 (29–490)                                        | 590 (320)                           | NQ                                 |
| Mediterranean Sea        | 30–45° N          | 2.5                                 | 10          | 0                      | 18 (12–28)                                          | 45 (21)                             | 0.2 (0.1–0.4)                      |
| Global                  |                   |                                     |             |                        |                                                     | 62 (52–73)                           | 137 (9.2)                          |

Earth Syst. Sci. Data, 4, 47–73, 2012  www.earth-syst-sci-data.net/4/47/2012/
than the peak of the distribution. The arithmetic mean in this basin is more than one order of magnitude higher than the geometric mean. In the North and South Pacific, the N\textsubscript{2} fixation rates only differ in 2 orders of magnitude, without any data lower than 10 \(\mu\text{mol N m}^{-2} \text{d}^{-1}\) (Fig. 10). Thus, the difference between the geometric and arithmetic means is smaller in the Pacific than in the North Atlantic. Although the geometric mean N\textsubscript{2} fixation rate in the North Atlantic is about one order of magnitude lower than those in the North and South Pacific, the peaks of the distributions are actually much closer in these basins (Fig. 10; also see Table 6). Compared to the North Atlantic, the Pacific is not intensively sampled for N\textsubscript{2} fixation. Thus, we need more samples with wider spatial coverage from the Pacific to better evaluate the global N\textsubscript{2} fixation rate.

Based on depth-integrated data, the global mean estimate of diazotrophic biomass from cell counts is 1–2 orders of magnitude lower than that from \(nifH\)-based abundances (Tables 7 and 8), which are mostly contributed by the dominant taxon, *Trichodesmium* (Table 5). But considering the log-normal distributions, difference of one or two orders of magnitude is relatively small compared to the span of the distribution especially when the number of samples is limited (see Fig. 7 as an example). More samples are needed in order to confidently validate the usage of \(nifH\) data for estimating diazotrophic abundances and the assumption of 1 : 1 conversion from \(nifH\) to cell densities.

The above global estimates are based on default values of the biomass conversion factors. In order to show the effects of the variation of the biomass conversion factors, we have applied the upper and lower bounds of these conversion factors for all diazotrophic subtypes in order to revise global diazotrophic biomass estimates following the same procedure described above. The estimated global geometric mean diazotrophic biomass can vary by about ± 70 % from default estimates, ranging in 0.8–3.5 Tg C (default 2.1 Tg C) based on cell count data, or in 28–170 Tg C (default 89 Tg C) based on \(nifH\)-based data (Table 2).
Figure 10. Histograms of depth-integrated total $N_2$ fixation rates (after binned on $3^\circ \times 3^\circ$ grid) in North Atlantic (blue), South Atlantic (green), North Pacific (red) and South Pacific (black). For each basin, filled diamond and empty circle mark the location of the geometric mean and the arithmetic mean, respectively.

4 Conclusions and recommendations for use

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

With the recent findings that most of the previous $N_2$ fixation rates have been underestimated when the $^{15}N_2$ tracer was added to the incubations as gas bubbles, it can be expected that higher $N_2$ fixation rates will be reported in the near future using the new method by adding $^{15}N_2$ in a dissolved form. Each data record needs to be well documented for the applied $^{15}N_2$ method, so that future users can identify which data records are likely underestimated.

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

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

Acknowledgements. Y.-W.L., S.C.D., D.B., M.J.C., D.M.K., P.H.M., K.A.T., A.E.W. and J.P.Z. are funded by C-MORE project (NSF EF-0424599). L.A.A. and D.J.M. gratefully acknowledge support of NSF OCE-0925284. M.B. acknowledges funding from projects CALBEX (CTM2007-66408-C02-02) and CYFOD (CTM2008-00915-E), and FPI fellowship (BES-2008-006985) by MICINN (Spain). S.B. is funded by IRD and data are contributions of the BOUM experiment and ISOfeRIX Project of the French national LEFE-CYBER program, the European IP SESAME and the international IMBER project. M.J.C., R.M.L. and D.M.K. acknowledge NSF OCE OCE-08-50827 for supporting $N_2$ fixation measurements at Station ALOHA. R.J.L. and J.L. are funded by the German BMBF project Sopran. D.M.K., P.H.M., K.A.T. and J.P.Z., are funded by the Gordon and Betty Moore Foundation. C.M.M. is funded by NERC, UK. B.M.-C. was supported by the Ramon y Cajal program from the Spanish Ministry of Science and Technology. A.P.R. received funding through NERC OCEANS2025. A.E.W. acknowledges funding from NSF OCE-0726422 (to F. Prahl, OSU). We thank Nicolas Gruber and another anonymous referee for their constructive reviews.

Edited by: S. Pesant

References

Arin, L., Berdalet, E., Marrasé, C., Estrada, M., Guixa-Boixereu, N., and Dolan, J.: Particulate DNA and protein relative to microorganism biomass and detritus in the Catalano-Balearic Sea (NW Mediterranean) during stratification, J. Plankton Res., 21, 1299–1316, 1999.

Bailiff, M. D. and Karl, D. M.: Dissolved and Particulate DNA Dynamics during a Spring Bloom in the Antarctic Peninsula Region,
Capone, D. G., Subramaniam, A., Montoya, J. P., Voss, M., Humberg, C., Johansen, A. M., Sievert, R. L., and Carpenter, E. J.: An extensive bloom of the N2-fixing cyanobacterium *Trichodesmium erythraeum* in the central Arabian Sea, Mar. Ecol.-Prog. Ser., 172, 281–292, doi:10.3335/meps172281, 1998.

Capone, D. G., Burns, J. A., Montoya, J. P., Subramaniam, A., Mahaffey, C., Gunderson, T., Michaels, A. F., and Carpenter, E. J.: Nitrogen fixation by *Trichodesmium* spp.: An important source of new nitrogen to the tropical and subtropical North Atlantic Ocean, Global Biogeochem. Cy., 19, GB2024, doi:10.1029/2004GB002331, 2005.

Carpenter, E. J.: Nitrogen fixation by marine *Oscillatoria (Trichodesmium)* in the world’s oceans, in: Nitrogen in the Marine Environment, edited by: Carpenter, E. J. and Capone, D. G., Academic Press, New York, 65–103, 1983.

Carpenter, E. J. and Capone, D. G.: Nitrogen Fixation in the Marine Environment, in: Nitrogen in the Marine Environment, edited by: Capone, D. G., Bronk, D. A., Mulholland, M. R., and Carpenter, E. J., Academic Press, London, 1197–1218, 2008.

Carpenter, E. J. and Price, C. C.: Nitrogen fixation, distribution, and production of *Oscillatoria (Trichodesmium)* spp. in the western Sargasso and Caribbean Seas, Limnol. Oceanogr., 22, 60–72, 1977.

Carpenter, E. J. and Romans, K.: Major role of the cyanobacterium *Trichodesmium* in nutrient cycling in the North Atlantic Ocean, Science, 254, 1356–1358, doi:10.1126/science.254.5036.1356, 1991.

Carpenter, E. J., Subramaniam, A., and Capone, D. G.: Biomass and primary productivity of the cyanobacterium, *Trichodesmium* spp., in the tropical North 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:10.3334/meps259047, 2003.

Chen, Y. L. L., Chen, H. Y., Tuo, S. H., and Ohki, K.: Seasonal dynamics of new production from *Trichodesmium* N2 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, 13–14, 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. Envir. Microbiol., 71, 5362–5370, doi:10.1128/Aem.71.9.5362-5370.2005, 2005b.

Church, M. J., Mahaffey, C., Letelier, R. M., Lukas, R., Zehr, J. P., and Karl, D. M.: Physical forcing of nitrogen fixation and diazotroph community structure in the North
Karl, D. M., Letelier, R., Tupas, L., Dore, J., Christian, J., and Karl, D., Michaels, A., Bergman, B., Capone, D., Carpenter, E., Hynes, A. M., Webb, E. A., Doney, S. C., and Waterbury, J. B.: Phytoplankton observations in the western Southwestern Sagasso Sea. J. Geol., 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.

Hulburt, E. M.: The distribution of Trichodesmium spp. in the productivity of the subtropical North Pacific Ocean. Mar. Ecol.-Prog. Ser., 133, 263–273, 2005.

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.

Hulburt, E. M.: Phytoplankton in the Southwestern 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., Webb, E. A., Doney, S. C., and Waterbury, J. B.: Comparison of cultured Trichodesmium (Cyanophyceae) with species characterized from the field. J. Phycol., 48, 196–210, doi:10.1111/j.1529-8817.2011.01096.x, 2012.

Janson, S., Wouters, J., Bergman, B., and Carpenter, E. J.: Host specificity in the Richelia-diatom symbiosis revealed by hetR gene sequence analysis. Environ. Microbiol., 1, 431–438, doi:10.1038/s41598-019-01645-3, 2019.

Koch, A. L.: Logarithm in Biology. I. 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.

Langlois, R. J., LaRoche, J., and Raab, P. A.: Diazotrophic diversity and distribution in the tropical and subtropical Atlantic ocean. Appl. Environ. Microbiol., 71, 7910–7919, doi:10.1128/Aem.71.12.7910-7919.2005.2005.

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

LaRoche, J. and Breitharth, 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.

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.

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.
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 N2 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 N2 fixation. Appl. Environ. Microbiol., 62, 986–993, 1996.

Montoya, J. P., Houll, C. M., Zehr, J. P., Hansen, A., Villareal, T. A., and Capone, D. G.: High rates of N2 fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean. Nature, 430, 1027–1032, doi:10.1038/nature02824, 2004.

Moore, C. M., Mills, M. M., Achterberg, E. P., Geider, R. J., LaRoche, J., Lucas, M. I., McDonagh, E. L., Pan, X., Poulton, A. J., Rijkenberg, M. J. A., Suggett, D. J., Ussher, S. J., and Woodward, E. M. S.: Large-scale distribution of Atlantic nitrogen fixation controlled by iron availability. Nat. Geosci., 2, 867–871, doi:10.1038/ngeo667, 2009.

Mourrato-Carballido, B., Grañ a, R., Fernández, A., Bode, A., Varela, M., Domínguez, J. F., Escá nez, J., Armas, D. D., and Marañón, E.: Importance of N2 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 rate of nitrogen fixed by diazotrophs in the ocean. Biogeosciences, 4, 37–51, doi:10.5194/bg-4-37-2007, 2007.

Mulholland, M. R., Bronk, D. A., and Capone, D. G.: Dinitrogen fixation and release of ammonium and dissolved organic nitrogen by Trichodesmium IMS101. Aquat. Microb. Ecol., 37, 85–94, doi:10.3354/ame037085, 2004.

Mulholland, M. R., Bernhardt, P. W., Blanco-Garcia, J. L., Mannino, A., Hyde, K., Mondragon, E., Turk, K., Moisander, P. H., and Zehr, J. P.: Rates of dinitrogen fixation and the abundance of diazotrophs in North American coastal waters between Cape Hatteras and Georges Bank, Limnol. Oceanogr., 57, 1067–1083, doi:10.4319/lo.2012.57.4.1067, 2012.

Needoba, J. A., Foster, R. A., Sakamoto, C., Zehr, J. P., and Johnson, K. S.: Nitrogen fixation by unicellular diazotrophic cyanobacteria in the temperate oligotrophic North Pacific Ocean, Limnol. Oceanogr., 52, 1317–1327, doi:10.4319/lo.2007.52.4.1317, 2007.

Orcutt, K. M., Lipschultz, F., Gundersen, K., Arimoto, R., Michaels, A. F., Knop, A. H., and Gallon, J. R.: A seasonal study of the significance of N2 fixation by Trichodesmium spp. at the Bermuda Atlantic Time-series Study (BATS) site, Deep-Sea Res. Pt. II, 48, 1583–1608, doi:10.1016/S0967-0645(00)00157-0, 2001.

Postgate, J. R.: The Fundamentals of Nitrogen Fixation, Cambridge University Press, New York, 1998.

Poulton, A. J., Stinchcombe, M. C., and Quartly, G. D.: High numbers of Trichodesmium and diazotrophic diatoms in the southwest Indian Ocean. Geophys. Res. Lett., 36, L15610, doi:10.1029/2009GL039719, 2009.

Rahav, E., Herut, B., Stambler, N., Bar-Zeev, E., Gertman, I., Mulholland, R. M., and Berman-Frank, I.: Summertime dinitrogen fixation in a gyre and eddy of the eastern Mediterranean Sea, Geophys. Res. Lett., in review, 2011.

Rahav, E., Herut, B., Levi, A., Mulholland, M., and Berman-Frank, I.: Controls on surface dinitrogen fixation across the Mediterranean during springtime, in preparation, 2012.

Rahav, E., Herut, B., Levi, A., Mulholland, M., and Berman-Frank, I.: Dinitrogen fixation at the stratified and mixed periods of the Gulf of Aqaba (Red Sea), in preparation, 2012b.

Raimbault, P. and Garcia, N.: Evidence for efficient regenerated production and dinitrogen fixation in nitrogen-deficient waters of the South Pacific Ocean: impact on new and export production estimates. Biogeosciences, 5, 323–338, doi:10.5194/bg-5-323-2008, 2008.

Reddy, K. J., Haskell, J. B., Sherman, D. M., and Sherman, L. A.: Unicellular, Aerobic Nitrogen-Fixing Cyanobacteria of the Genus Cyanothece. J. Bacteriol., 175, 1284–1292, 1993.

Rees, A. P., Law, C. S., and Woodward, E. M. S.: High rates of nitrogen fixation during an in-situ phosphate release experiment in the Eastern Mediterranean Sea, Geophys. Res. Lett., 33, L10067, doi:10.1029/2006GL025791, 2006.

Rees, A. P., Gilbert, J. A., and Kelly-Gerrey, 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 Eskimazi, E.: On the mechanism of red tide of Trichodesmium in Recife, northwest Brazil, with some considerations of the relation to the human disease “Tamadare fever”, Trab. Inst. Oceanogr. Univ. Recife, 5, 7–49, 1966.

Scranton, M. I.: Hydrogen cycling in the waters near Bermuda: the role of the nitrogen fixer, Oscillatoria thiebuatii, Deep-Sea Res., 31, 133–143, 1984.

Shiozaki, T., Furuya, K., Kodama, T., Kitajima, S., Takeda, S., Takeamura, T., and Kanda, J.: New estimation of N2 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 Ocean. Environ. Microbiol., 14, 641–654, doi:10.1111/j.1462-2920.2011.02606.x, 2012.
Tuit, C., Waterbury, J., and Ravizzaz, G.: Diel variation of molybdenum and iron in marine diazotrophic cyanobacteria, Limnol. Oceanogr., 49, 978–990, doi:10.4319/lo.2004.49.4.0978, 2004.

Turk, K. A., Rees, A. P., Zehr, J. P., Pereira, N., Swift, P., Shelley, R., Lohan, M., Woodward, E. M. S., and Gilbert, J.: Nitrogen fixation and nitrogenase (nifH) expression in tropical waters of the eastern North Atlantic, ISME J., 5, 1201–1212, doi:10.1038/ismej.2010.205, 2011.

Tyrrell, T.: The relative influences of nitrogen and phosphorus on oceanic primary production, Nature, 400, 525–531, doi:10.1038/22941, 1999.

Tyrrell, T., Marañón, E., Poulton, A. J., Bowie, A. R., Harbour, D. S., and Woodward, E. M. S.: Large-scale latitudinal distribution of Trichodesmium spp. in the Atlantic Ocean, J. Plankton Res., 25, 405–416, doi:10.1093/plankt/25.4.405, 2003.

Venrick, E. L.: Distribution and Significance of Richelia Intracellularis Schmidt in North Pacific Central Gyre, Limnol. Oceanogr., 19, 437–445, 1974.

Verity, P. G., Robertson, C. Y., Trono, 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:10.1080/0007161890065374, 1989.

Villareal, T. A.: Laboratory Culture and Preliminary Characterization of the Nitrogen-Fixing Rhizosolenia-Richelia Symbiosis, Mar. Ecol., 11, 117–132, doi:10.1111/j.1439-0485.1990.tb00233.x, 1990.

Villareal, T. A.: Marine nitrogen-fixing diatom-cyanobacteria symbioses, in: Marine Pelagic Cyanobacteria: Trichodesmium and other diazotrophs, edited by: Carpenter, E. J., Capone, D. G., and Rueter, J. G., NATO ASI Series, 362, Kluwer Academic Publishers, Boston, 163–175, 1992.

Villareal, T. A., Adornato, L., Wilson, C., and Schoenbaechler, C. A.: Summer blooms of diatom-diazotroph assemblages and surface chlorophyll in the North Pacific gyre: A disconnect, J. Geophys. Res.-Oceans, 116, C03001, doi:10.1029/2010jc006268, 2011.

Watkins-Brandt, K. S., Letelier, R. M., Spitz, Y. H., Church, M. J., Böttjer, D., and White, A. E.: Addition of inorganic or organic phosphorus enhances nitrogen fixation in the oligotrophic North Pacific, Mar. Ecol.-Prog. Ser., 432, 17–29, doi:10.3354/meps09147, 2011.

Webb, E. A., Ehrenreich, I. M., Brown, S. L., Valois, F. W., and Waterbury, J. B.: Phenotypic and genotypic characterization of multiple strains of the diazotrophic cyanobacterium, 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, 2012.

Wilson, S. T., Böttjer, D., Church, M. J., and Karl, D. M.: Comparative assessment of nitrogen fixation methodologies conducted in the oligotrophic North Pacific Ocean, Appl. Environ. Microbiol., 78, 6516–6523, doi:10.1128/AEM.01146-12, 2012.

Winn, C. D. and Karl, D. M.: Diel Nucleic-Acid Synthesis and Particulate DNA Concentrations – Conflicts with Division Rate Estimates by DNA Accumulation, Limnol. Oceanogr., 31, 637–645, 1986.

Wu, J. F., Chung, S. W., Wen, L. S., Liu, K. K., Chen, Y. L. L., Chen, H. Y., and Karl, D. M.: Dissolved inorganic phosphorus, dissolved iron, and Trichodesmium in the oligotrophic South China Sea, Global Biogeochem. Cy., 17, 1008, doi:10.1029/2002gb001924, 2003.

Yoge, T., Rahav, E., Bar-Zeev, E., Man-Aharonovich, D., Stambler, N., Kress, N., Beja, O., Mulholland, M. R., Herut, B., and Berman-Frank, I.: Is dinitrogen fixation significant in the Levantine Basin, East Mediterranean Sea?, Environ. Microbiol., 13, 854–871, doi:10.1111/j.1462-2920.2010.02402.x, 2011.

Zehr, J. P.: Nitrogen fixation by marine cyanobacteria, Trends Microbiol., 19, 162–173, doi:10.1016/j.tim.2010.12.004, 2011.

Zehr, J. P., Mellon, M. T., and Zani, S.: New nitrogen-fixing microorganisms detected in oligotrophic oceans by amplification of nitrogenase (nifH) genes, Appl. Environ. Microbiol., 64, 3444–3450, 1998.

Zehr, J. P., Waterbury, J. B., Turner, P. J., Montoya, J. P., Omoregie, E., Steward, G. F., Hansen, A., and Karl, D. M.: Unicellular cyanobacteria fix N2 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 N2-Fixing Cyanobacteria Lack Oxygenic Photosystem II, Science, 322, 1110–1112, doi:10.1126/science.1165340, 2008.