Geophysical Research Letters

RESEARCH LETTER

10.1029/2019GL084376

Key Points:

- Based on a literature review, the database of marine diazotrophs is updated, more than doubling the number of observations
- The distributions of four major diazotrophic groups distinctly relate to temperature, light and nutrients at the global scale
- A machine learning method is applied to derive global maps of these diazotrophic groups in the global ocean

Supporting Information:

- Supporting Information S1
- Table S1
- Movie S1
- Movie S2
- Movie S3
- Movie S4

Correspondence to:

W. Tang,
weiyi.tang@duke.edu

Citation:

Tang, W., & Cassar, N. (2019). Data-driven modeling of the distribution of diazotrophs in the global ocean. *Geophysical Research Letters*, 46, 12,258–12,269. <https://doi.org/10.1029/2019GL084376>

Received 1 JUL 2019

Accepted 29 OCT 2019

Accepted article online 9 NOV 2019

Published online 13 NOV 2019

Data-Driven Modeling of the Distribution of Diazotrophs in the Global Ocean

Weiyi Tang^{1,2} and Nicolas Cassar^{1,3}

¹Division of Earth and Ocean Sciences, Nicholas School of the Environment, Duke University, Durham, NC, USA, ²Now at Department of Geosciences, Princeton University, Princeton, NJ, USA, ³Laboratoire des Sciences de l'Environnement Marin, UMR 6539 UBO/CNRS/IRD/IFREMER, Institut Universitaire Européen de la Mer, Brest, France

Abstract Diazotrophs play a critical role in the biogeochemical cycling of nitrogen, carbon, and other elements in the global ocean. Despite their well-recognized role, the diversity, abundance, and distribution of diazotrophs in the world's ocean remain poorly characterized largely due to limited observations. Here we update the database of diazotroph *nifH* gene abundances and assess how environmental factors may regulate diazotrophs at the global scale. Our meta-analysis more than doubles the number of observations in the previous database. Using linear and nonlinear regressions, we find that the abundances of *Trichodesmium*, UCYN-A, UCYN-B, and *Richelia* relate differently to temperature, light, and nutrients. We further apply a random forest algorithm to estimate the global distributions of these diazotrophic groups, identifying undersampled potential hot spots of diazotrophy in the South Atlantic and southern Indian Ocean, and in coastal waters. The distinct ecophysiologicals of diazotrophs highlighted here argue for separate parameterizations of different diazotrophs in model simulations.

Plain Language Summary Microbial communities drive the cycling of critical elements like carbon and nitrogen in the ocean. By converting N₂ into more bioavailable nitrogen, diazotrophs alleviate nitrogen limitation and support primary production. Despite their importance, their distributions are poorly characterized in great part due to limited observations. Here we compile from the literature observations to update the global database of marine diazotrophs. We also assess how the abundance and distribution of different types of diazotrophs at the global scale relate to environmental factors, including temperature, depth, and nutrients. Finally, we use a random forest machine learning method to predict the distribution of different types of diazotrophs in the world's ocean. Our results highlight the need for observations over broader oceanic regimes and a more granular representation of diazotrophy in models.

1. Introduction

Marine microbes drive biogeochemical processes important to our global climate (Falkowski et al., 2008; Sunagawa et al., 2015; Worden et al., 2015). Of particular note are diazotrophs: key microorganisms converting N₂ gas into bioavailable nitrogen, thereby relieving nitrogen limitation and supporting marine production in many regions of the world's oceans (Karl et al., 1997; Sohm et al., 2011). *Trichodesmium* and diazotrophs associated with diatoms (e.g., *Richelia*) have traditionally been assumed to be the dominant diazotrophs in the ocean (Capone et al., 1997; Villareal, 1991). However, thanks to advances in molecular techniques and gene sequencing, our appreciation for diazotroph diversity is rapidly evolving (Delmont et al., 2018; Zehr, 2011; Zehr et al., 1998). For example, unicellular cyanobacterial diazotrophs have recently been found to account for a large proportion of the diazotrophic community and to contribute substantially to N₂ fixation in the tropical Atlantic (Martínez-Pérez et al., 2016). Despite much progress, our current view of the global distribution of marine diazotrophs remains qualitative.

Luo et al. (2012) compiled the first databases of diazotrophs and N₂ fixation in the global ocean. While their database of N₂ fixation rates has been the subject of several analyses (Luo et al., 2014; Tang, Li, & Cassar, 2019), their database on the distribution of diazotrophs has received less attention. Luo et al. (2012) assembled 109 and 921 observations of depth-integrated and volumetric diazotroph abundances, respectively, as estimated by quantitative polymerase chain reaction (qPCR) assays targeting the *nifH* genes of various diazotrophs including *Trichodesmium*, UCYN-A1, UCYN-A2, UCYN-B, UCYN-C, *Richelia*, *Calothrix*, and some noncyanobacterial diazotrophs. Since publication of the database, the number of observations has rapidly expanded, with new data collected in coastal, aphotic, and polar waters (Benavides et al.,

2018; Shiozaki, Fujiwara, et al., 2018; Tang, Wang, et al., 2019). In light of such progress, updating the Luo et al. (2012) database of diazotroph abundances and reevaluating the environmental controls on their distributions in the world's oceans represents a timely effort.

Prognostic parameterizations have been used to model the regional and global distribution of diazotrophs. For example, Hood et al. (2001) and Fennel et al. (2001) simulated the seasonal cycles of *Trichodesmium* at BATS and ALOHA using 1-D biogeochemical models, respectively. More recently, Paulsen et al. (2017) included a prognostic representation of *Trichodesmium* in a global ocean biogeochemical model. In addition, resource competition theory and nutrient supply ratios have also been applied to define the biogeography of diazotrophs (Follett et al., 2018; Ward et al., 2013). Most models however focus on *Trichodesmium*. As a result, the biogeography of diazotrophs with ecophysologies distinct from *Trichodesmium* is poorly constrained. As recently discussed in Tang, Li, and Cassar (2019), a model, which incorporates other dominant diazotrophic groups, may improve the characterization of the distribution of N_2 fixation.

In this study, we present an initial attempt at constructing data-driven biogeographies of various diazotrophs. To that end, we first update the meta-analysis of Luo et al. (2012) on diazotroph abundances. After identifying the dominant environmental predictors of diazotroph abundances, we apply a random forest (RF) machine learning technique to simulate the global distributions of the four major marine diazotrophs *Trichodesmium*, UCYN-A, UCYN-B (i.e., *Crocospaera*), and *Richelia*. We conclude by comparing the geographical distributions of our data-derived estimates to trait-based model simulations.

2. Materials and Methods

2.1. Updating the Global Diazotrophs Database

We supplemented the dataset presented in Luo et al. (2012) with measurements of *nifH* gene abundances using qPCR from 17 additional publications (Berthelot et al., 2017; Halm et al., 2012; Harding et al., 2018; Hashimoto et al., 2016; Henke et al., 2018; Krupke et al., 2013; Martínez-Pérez et al., 2016; Moreira-Coello et al., 2017; Shiozaki, Chen, et al., 2014; Shiozaki, Ijichi, et al., 2014; Shiozaki et al., 2017, 2015; Shiozaki, Bombar, et al., 2018; Shiozaki, Fujiwara, et al., 2018; Shiozaki, Kondo, et al., 2018; Stenegren et al., 2018; Wen et al., 2017). These data were either extracted from the published papers or provided by the authors. One hundred fifty-four and 1,085 new observations of depth-integrated and volumetric diazotroph abundances (i.e., *nifH* gene copies) were obtained, respectively. This addition represents a 141% and a 118% increase in the number of depth-integrated and volumetric data points compared to the database of Luo et al. (2012). The depth-integrated diazotroph abundances (*nifH* copies m^{-2}) were calculated by integrating volumetric diazotroph abundances (*nifH* copies m^{-3}) over the euphotic zone or from certain depths to the surface using the trapezoidal rule, as in Luo et al. (2012). Our new database includes 223 (1,428), 253 (1,575), 226 (1,424), and 144 (958) observations of depth-integrated (volumetric) abundances of *Trichodesmium*, UCYN-A, UCYN-B, and *Richelia*, respectively, as shown in Table S1 in the supporting information (<https://doi.org/10.1594/PANGAEA.905108>). The updated global distributions of the four major diazotrophs are shown in Figure 1, with the vertical profiles of their volumetric abundances presented in Figures S1–S4 in the supporting information. The following analyses are based on *nifH* gene copy numbers. Because of variations in *nifH* gene copy number per cell in different diazotrophs and methodological issues (e.g., nonspecificity of primers and probes used in qPCR), conversion to cell numbers is not straightforward. More details can be found in Text S1 in the supporting information (see also White et al., 2018).

2.2. Correlation Analyses and Machine Learning Simulation of Diazotrophs

Correlation analyses (simple linear regression, polynomial regression, partial regression, and redundancy analysis) were conducted between the volumetric abundances of diazotrophs and field-measured environmental variables hypothesized to regulate diazotrophy, including depth, temperature, and nutrient concentrations (Karl et al., 2002; Sohm et al., 2011). Conversely, based on environmental descriptors in Luo et al. (2014), the depth-integrated abundances were matched and compared to various contemporaneous environmental factors derived from satellite observations, reanalyses and models (Table S2), including solar radiation (I_s), wind speed (WS), sea surface temperature (SST), sea surface salinity (SSS), surface dissolved inorganic nitrate (DIN), surface dissolved inorganic phosphate (DIP), surface excess phosphorus ($P^* = DIP - DIN/16$), minimum oxygen in the upper 500 m (DO_{min}), photosynthetically

Figure 1. Global distributions of the depth-integrated *nifH* gene abundances of four major diazotrophs (a) *Trichodesmium*, (b) UCYN-A, (c) UCYN-B, and (d) *Richelia*.

available radiation (PAR), mixed layer depth (MLD), averaged PAR in the mixed layer (PAR_{mld}), chlorophyll-*a* concentration ([Chl]), and modeled surface iron concentration (Fe). Overall, 193, 218, 198, and 131 matches to contemporaneous environmental variables were obtained for the depth-integrated *Trichodesmium*, UCYN-A, UCYN-B, and *Richelia* abundance data, respectively. We first applied simple regressions to evaluate the correlations between depth-integrated diazotroph abundances and environmental factors. We then used a RF algorithm to simulate the global depth-integrated distribution of the four diazotroph groups based on the environmental variables described above (algorithms are available in <https://doi.org/10.1594/PANGAEA.905108>). In addition, we applied RF to estimate volumetric diazotroph abundances but only for surface volumetric diazotroph abundances (Figure S5). This is because of the lack of field-measured environmental variables matched to volumetric diazotroph abundances, especially for the observations at depth. For example, only 55 locations have concurrent surface observations of the explanatory variables used in the model construction (temperature, salinity, nitrate, and phosphate) and volumetric *Trichodesmium* abundance. The processes of data matching and correlation analyses are shown in a workflow chart (Figure S6) and in Text S2 in the supporting information (Boyer et al., 2013; de Boyer Montégut et al., 2004; Kalnay et al., 1996; Luo et al., 2014; Moore et al., 2013; Tang, Li, & Cassar, 2019; Taylor et al., 2012).

Random forest builds an ensemble of decisions trees based on the bootstrap sampling and random selection of explanatory variables to improve the prediction accuracy (Breiman, 2001). This approach is increasingly applied in earth sciences along with other machine learning techniques (Bergen et al., 2019). For the construction of the RF model, we randomly divided our matches into a training-validation dataset (70%) and a testing dataset (30%). following other studies in environmental sciences (Gregor et al., 2017). In the model training, the most important parameters to define are the number of trees to grow and the number of random variables to choose at each split (Liaw & Wiener, 2002). These parameters were optimized by minimizing the out-of-bag mean square error. Five hundred trees were built, and six random variables were chosen in this study to minimize and stabilize the mean square error (Figure S7). The trained models were evaluated against the test datasets. After model construction and validation, the algorithm was used to simulate the

Figure 2. Relations between volumetric diazotroph abundances and contemporaneously field-observed environmental properties including depth, temperature, dissolved inorganic nitrogen (DIN), dissolved inorganic phosphorus (DIP) and the N:P ratio (DIN:DIP) color-coded by the density of observations (Eilers & Goeman, 2004). The full-range of observations are shown in Figure S11.

monthly abundance of the four diazotroph groups in the global ocean with a $2^{\circ} \times 2^{\circ}$ resolution. A thorough description of the RF method and its applications can be found in Breiman, 2001.

3. Results and Discussion

3.1. Distribution of Diazotrophs and Environmental Controls

In the updated database, most observations of marine diazotroph abundances come from the tropical Atlantic and western tropical Pacific (Figure 1), with sparse observations in the Indian Ocean and eastern Pacific. However, observations now extend to temperate and polar regions with the discovery of N_2 fixation in cold and nutrients-enriched environments (Harding et al., 2018; Shiozaki et al., 2017). *Trichodesmium* reaches up to 10^{11} copies m^{-2} near the warm and nutrient-poor Caribbean and Solomon Seas. This is likely due to *Trichodesmium*'s preference for higher temperature and lower DIN (Figure 2). Lab culture experiments have shown that *Trichodesmium* optimally grows at $\sim 26^{\circ}C$ (Breitbarth et al., 2007; Fu et al., 2014). UCYN-B displays a distribution similar to *Trichodesmium* in the Pacific and appears to be less abundant

in the North Atlantic (Figure 1). This pattern may be driven by UCYN-B's positive correlation to temperature (Moisander et al., 2010) and higher requirement for DIP (Figure 2). Therefore, the phosphate-depleted subtropical North Atlantic (Mather et al., 2008; Wu et al., 2000) may restrict the growth of UCYN-B. The diatom-associated diazotroph *Richelia* generally displays lower depth-integrated abundances (below 10^9 copies m^{-2}) but thrives in the warm and silicate-rich Amazon River plume (Foster et al., 2007; Weber et al., 2016). In contrast, UCYN-A presents a distribution distinct from the other diazotrophs (Stenegren et al., 2018). For example, UCYN-A has recently been shown to actively fix N_2 in the cold waters of the Arctic Ocean (Harding et al., 2018). Compared to other diazotrophic groups, UCYN-A has a somewhat lower optimal temperature at around 24°C. It is the only diazotroph analyzed in this study to be present at high abundance when temperatures are below 10°C (Figure 2).

The meta-analysis at the global scale also reveals some intriguing patterns not necessarily captured by regional studies. For example, previous studies have attributed depth niches to diazotrophs, with *Trichodesmium* and UCYN-A generally believed to dominate at the surface and subsurface, respectively (Stenegren et al., 2018). However, no clear separation of diazotrophic groups with depth is found at the global scale because all groups in this study generally decrease with depth (Figure 2; $r < 0$ and $p < 0.01$ for all diazotrophs) with UCYN-A and UCYN-B extending deeper in the water column (Figure S8). However, diazotrophic blooms occasionally form at the subsurface. For example, up to 10^5 copies L^{-1} of *Trichodesmium* and UCYN-A are observed down to depths of 100 m.

As mentioned above, temperature seems to be an important predictor as shown in the redundancy analysis where *Trichodesmium*, UCYN-B, and *Richelia* cluster together and their variations follow the changes in temperature (Figure S9). The upper bound temperatures are consistent with culture experiments. For example, the optimal temperature of UCYN-B is $\sim 30^\circ\text{C}$ (Fu et al., 2014), in line with the temperature where the maximum abundance of UCYN-B is observed in the field (Figure S10). Temperature has been shown to set upper bounds on other physiological rates including the maximum growth rate of phytoplankton (Eppley, 1972). The ability of diazotrophs to reach maximum abundance at a given temperature is likely modulated by other growth factors, including light availability as a function of depth, nitrate, phosphate, top-down controls (Wang et al., 2019), and other yet-to-be determined factors (Figure S12). This may in part explain the different temperature-diazotroph relationships observed at regional scales (Church et al., 2009; Shiozaki, Chen, et al., 2014). However, our estimated temperature upper bounds may not capture the true full range (e.g., limited observations of UCYN-A from 5 to 10°C). Our meta-analysis also substantiates that diazotrophs may not be restricted to nitrogen-poor waters. Although diazotrophs generally become less abundant with increasing DIN in response to competitive exclusion (Ward et al., 2013), they are still present, e.g., 10^4 copies L^{-1} above 5 μM DIN (Figure S11). This is consistent with lab cultures showing substantial N_2 fixation rates (Knapp, 2012) and with in situ measurements of diazotrophs growth in the presence of high DIN (Turk-Kubo et al., 2018).

Because environmental properties are correlated to each other, deconvolving their independent effect on the distribution of diazotrophs is not trivial. For example, our analyses show that the diazotroph abundances are at first-order negatively correlated to phosphate concentration ($r < 0$, $p < 0.01$ for all diazotrophs except *Richelia* with p value at 0.16) shown in Figures 2 and S11. This is surprising because of the important diazotrophic requirement for phosphorus (Sañudo-Wilhelmy et al., 2001; Sohm et al., 2008). This counterintuitive pattern could be due to the correlation of phosphate to other factors negatively impacting the diazotrophs, including lower temperatures and light availability with depth (Figure S12). As expected, DIP correlates with depth (Figure S9). When only considering data collected at the surface (depth ≤ 5 m), the negative correlations weaken dramatically, notably for UCYN-B (correlation coefficient changing from -0.24 to 0.23). After removing the effects of depth and temperature using partial regression analyses, the correlation between diazotroph abundances (except for UCYN-A) and DIP becomes nonsignificant. Diazotroph abundances also display an ambiguous relationship to the N:P ratio and iron. Because iron is a key element in the nitrogenase enzyme (Hoffman et al., 2014), it is believed to be a dominant control on the distribution of diazotrophs when it is limiting (Moore et al., 2009). Simultaneous field observations of iron and diazotrophs are limited, making it difficult to evaluate the effect of iron on their global distribution (Figure S11). We also do not find a clear relationship between depth-integrated diazotroph abundances and modeled Fe concentrations (Figures S13–S16). In general, the relations between depth-integrated diazotrophs and matched

contemporaneous satellite-parameters (Figures S13–S116) support the relations shown between volumetric diazotrophs abundances and field-measured environmental variables (Figures 2 and S11).

3.2. Machine Learning Estimates and Comparison to Other Model Simulations

We use the RF machine learning algorithm to estimate the depth-integrated abundance of diazotrophs based on the environmental factors described in Table S2. The most important predictors chosen during model construction vary substantially among diazotrophic groups (Figure S17), further highlighting the distinct biogeochemical controls on diazotrophs at the global scale. SST is the most importance predictor of *Trichodesmium* and UCYN-B, in agreement with our previous correlation analyses between temperature and diazotroph abundances. Our RF estimates are in line with observations but produce significant scatter (Figure S18). In some cases, our algorithm produces false positives, incorrectly predicting the presence of diazotrophs when they are not observed in the original dataset. This could result from some controls on diazotroph abundance not being included in our predictors. For example, Wang et al. (2019) recently suggested that grazing, not accounted for in our RF algorithm, may regulate the abundance of diazotrophs. Below, we describe and compare our estimates with other model simulations at global and basin scales.

Global Ocean: Our RF algorithm predicts distinct biogeographies or niches for the four major diazotroph groups despite their overlap in the subtropical Pacific, western tropical Atlantic near the Caribbean Sea and subtropical Indian Ocean (Figure 3e). *Trichodesmium* is prevalent in the tropical and subtropical ocean within the approximate range of 30°N and 30°S, dominating in the tropical Atlantic, western Indian Ocean, and western Pacific. In contrast, the UCYN-A distribution extends to temperate and polar regions. UCYN-B shares a similar distribution as *Trichodesmium* but exhibits smaller depth-integrated abundances. *Richelia* shows a relatively homogeneous distribution in the tropical ocean with hot spots predicted in the Amazon River plume, near Hawaii, and in the Mediterranean Sea, in agreement with observations from studies not included in our training dataset (Crombet et al., 2011; Villareal et al., 2012; Weber et al., 2016). In addition, although the quantification of diazotrophs in coastal oceans is scarce in our dataset, diazotrophic hot spots are predicted in coastal waters (e.g., eastern American coast, coast of islands in the western Pacific), in line with the recent studies of Tang, Wang, et al. (2019) and Mulholland et al. (2019). More broadly, the RF-estimated distributions of diazotrophs combining *Trichodesmium*, UCYN-A, UCYN-B, and *Richelia* match the modeled distribution of N₂ fixation by Tang, Li, and Cassar (2019). To the best of our knowledge, only one other study has estimated the distribution of diazotrophs other than *Trichodesmium* (Monteiro et al., 2010). In their study, they estimated the coexistence of analogs of *Trichodesmium*, unicellular diazotrophs and diatom-diazotroph associations.

Atlantic: The highest abundance of *Trichodesmium* has been reported in the western tropical Atlantic (Capone et al., 2005). This pattern is consistently captured by our RF model and other simulations (Hood et al., 2004; Monteiro et al., 2010; Paulsen et al., 2017). While data are limited in the Gulf of Guinea off Africa, the RF-model also simulates a hot spot of *Trichodesmium* abundance in this region. Our RF also predicts seasonal cycles of *Trichodesmium* in the tropical Atlantic comparable to the model simulations of Hood et al. (2004) (see supplementary videos). A predicted shift from *Trichodesmium* in the western tropical Atlantic to UCYN-A in the eastern tropical Atlantic matches field observations (Benavides et al., 2016; Martínez-Pérez et al., 2016). On the other hand, the simulated high abundance of UCYN-A south of 30°S requires further confirmation as only a few observations are currently available. The RF-predicted lower abundance of UCYN-B in the North Atlantic matches field observations as described in section 3.1 (Figure 1).

Pacific: The abundance of diazotrophs is low at the equator and high in the flanking subtropical waters (Figure 3). This pattern may be driven by iron limitation near the equator especially in the eastern tropical Pacific (Ellwood et al., 2018; Knapp et al., 2016). Iron limitation is relieved in the western subtropical Pacific, notably close to islands (Shiozaki, Kodama, & Furuya, 2014). Our RF-estimate of the abundance of *Trichodesmium* matches a recent model based on the physiology of *Trichodesmium* (Dutheil et al., 2018). On the other hand, Paulsen et al. (2017) shows high abundance of *Trichodesmium* extending to the eastern Pacific where *Trichodesmium* blooms are frequently detected via satellite (Westberry & Siegel, 2006). Also worth noting is a discrepancy in the meridional distribution of diazotrophs as determined from different approaches. The model of Monteiro et al. (2010) displays substantially higher diazotroph abundances in the North Pacific than in the South Pacific similar to the diazotroph biogeography constrained by the Fe:

Figure 3. Simulated annual average abundance of (a) *Trichodesmium*, (b) UCYN-A, (c) UCYN-B, (d) *Richelia*, and (e) diazotrophs niches in the global ocean. The overlap regions are defined as regions where the abundances of all four diazotrophs exceed 10^6 *nifH* gene copies m^{-2} . A particular diazotrophic group is color-coded in regions where it is predicted to be twice more abundant than all the other groups.

N supply ratio (Ward et al., 2013), but differing from Paulsen et al. (2017), Dutheil et al. (2018), and our RF model.

Indian: Few studies have explored diazotrophy in the Indian Ocean (Figure 1). Although models generally simulate vigorous diazotrophy in this ocean basin, there are large discrepancies in the spatial distribution. Our RF-model predicts high abundance of diazotrophs in the tropical and subtropical waters near Madagascar and off the coast of western Australia, which is consistent with field observations of *Trichodesmium* and unicellular diazotrophs blooms in these areas (Holl et al., 2007; Poulton et al., 2009; Srokosz & Quartly, 2013). In comparison, Monteiro et al. (2010) and Paulsen et al. (2017) simulate abundant diazotrophs in the Arabian Sea and the Bay of Bengal, which is supported by some studies (Jyothibabu et al., 2017). While noncyanobacterial diazotrophy is currently not simulated in any of the models, recent studies

Figure 4. Conceptual diagram of an alternative approach to model marine N_2 fixation accounting for granularity in the ecophysiology of diazotrophs. (a and b) Distinct environmental controls among diazotrophs are coupled with (c) the distributions of environmental properties to simulate (d) global marine N_2 fixation. (b) Range and typical growth conditions for marine diazotrophs' growth based on culture experiments. T: temperature range (optimal temperature in the parentheses), $^{\circ}C$; light: light saturation intensity or half saturation constant, $\mu mol\ quanta\ m^{-2}\ s^{-1}$; Ks (P): half saturation constant of phosphate, μM ; Ks (Fe): half saturation constant of iron, nM. 1. Breitbarth et al. (2007); 2. Fu et al. (2014); 3. Breitbarth et al. (2008). Light saturation intensity; 4. Fu et al. (2005); 5. Boatman et al. (2018). Free iron concentration; 6. Garcia et al. (2013). Half saturation constant; 7. Jacq et al. (2014). Dissolved iron concentration; 8. Pyle (2011). Light saturation intensity for *Hemiaulus haukii*; 9. Follett et al. (2018).

have highlighted their potentially overlooked importance in the Indian Ocean (Bird & Wyman, 2013; Shiozaki, Ijichi, et al., 2014; Wu et al., 2019).

Undersampled potential hot spots of diazotrophy: Our RF-algorithm may be used to guide future sampling strategies. Most models generally restrict diazotrophs to the oligotrophic tropical and subtropical surface waters. However, this classic paradigm is increasingly being challenged by observations of diazotrophs in aphotic, polar, and coastal waters (Bonnet et al., 2013; Harding et al., 2018; Tang, Wang, et al., 2019). Our RF model indeed predicts high abundance of diazotrophs in the undersampled coastal waters, South Atlantic, southern Indian Ocean, and polar regions. Model prediction of the presence of UCYN-A at higher latitudes results in part from UCYN-A's lower optimal temperature and its tolerance of high DIN (Figure 2). While this is speculative as we are extrapolating beyond the ranges of some environmental factors used in the model training, our predictions are supported by recent reports of UCYN-A in Arctic waters (Harding et al., 2018; Shiozaki, Fujiwara, et al., 2018).

3.3. A More Granular Representation of Diazotrophs in Earth System Models

In light of the diverse ecophysiology of the broad spectrum of organisms fixing N_2 (Figures 4a and 4b), integration of multiple diazotrophic groups in addition to the most commonly represented *Trichodesmium* in

earth system models would improve our estimates of diazotroph distributions and N_2 fixation rates. One strategy to model N_2 fixation rates may be to combine their individual distributions and diazotrophic activities to represent global distribution of N_2 fixation (Figure 4d).

4. Conclusions

Despite a number of caveats, limitations, and uncertainties in our simulations, including varying ecotypes within a single group, PCR primer biases, varying *nifH* gene copy numbers per cell, limited and biased range of biomes sampled, and temporal mismatch between environmental forcing and diazotroph abundances as shown in the supporting information (Delmont et al., 2018; Farnelid et al., 2016; Sargent et al., 2016; Thompson et al., 2014; White et al., 2018), this study highlights directions for future field work by identifying regions worthy of further investigation because they may harbor diazotrophic hot spots (e.g., coastal waters; western tropical Atlantic), remain presently undersampled (e.g., South Atlantic, southern Indian Ocean), or produce large discrepancies in model simulations (e.g., polar regions). Our study also shows that four major diazotrophic groups display distinct biogeographic niches. Other diazotrophs not studied here such as non-cyanobacterial diazotrophs likely further diversify the range of marine ecosystems where N_2 fixation may be prevalent. With a changing ocean, the distributions of these groups are expected to evolve differently. A more granular representation of diazotrophs in earth system models may therefore improve our projections of feedbacks between marine microbial communities and the biogeochemical cycles of carbon and nitrogen.

Acknowledgments

We would like to thank Yawei Luo and colleagues for compiling the first global database of diazotrophs. We also would like to thank Takuhei Shiozaki (The University of Tokyo), Dalin Shi (Xiamen University), and Zuozhu Wen (Xiamen University) for providing access to their valuable diazotroph datasets. We would like to acknowledge CMIP5, NASA Ocean Color, NCEP/NCAR reanalysis, World Ocean Atlas, and Ifremer for processing and distributing their products. The updated dataset of diazotrophs, machine learning algorithms, codes, and model outputs have been deposited in PANGAEA (<https://doi.org/10.1594/PANGAEA.905108>). W. T. and N. C. are supported by an NSF-CAREER grant (1350710). N. C. is also supported by the "Laboratoire d'Excellence" LabexMER (ANR-10-LABX-19) and cofunded by a grant from the French government under the program "Investissements d'Avenir."

References

- Benavides, M., Moisaner, P. H., Daley, M. C., Bode, A., & Aristegui, J. (2016). Longitudinal variability of diazotroph abundances in the subtropical North Atlantic Ocean. *Journal of Plankton Research*, 38(3), 662–672. <https://doi.org/10.1093/plankt/fbv121>
- Benavides, M., Shoemaker, K. M., Moisaner, P. H., Niggemann, J., Dittmar, T., Duhamel, S., et al. (2018). Aphotic N_2 fixation along an oligotrophic to ultraoligotrophic transect in the western tropical South Pacific Ocean. *Biogeosciences*, 15(10), 3107–3119. <https://doi.org/10.5194/bg-15-3107-2018>
- Bergen, K. J., Johnson, P. A., de Hoop, M. V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. *Science*, 363(6433), eaau0323. <https://doi.org/10.1126/science.aau0323>
- Berthelot, H., Benavides, M., Moisaner, P. H., Grosso, O., & Bonnet, S. (2017). High-nitrogen fixation rates in the particulate and dissolved pools in the Western Tropical Pacific (Solomon and Bismarck Seas). *Geophysical Research Letters*, 44, 8414–8423. <https://doi.org/10.1002/2017GL073856>
- Bird, C., & Wyman, M. (2013). Transcriptionally active heterotrophic diazotrophs are widespread in the upper water column of the Arabian Sea. *FEMS Microbiology Ecology*, 84(1), 189–200. <https://doi.org/10.1111/1574-6941.12049>
- Boatman, T. G., Oxborough, K., Gledhill, M., Lawson, T., & Geider, R. J. (2018). An integrated response of *Trichodesmium erythraeum* IMS101 growth and photo-physiology to iron, CO_2 , and light intensity. *Frontiers in Microbiology*, 9, 624. <https://doi.org/10.3389/fmicb.2018.00624>
- Bonnet, S., Dekaezemacker, J., Turk-Kubo, K. A., Moutin, T., Hamersley, R. M., Grosso, O., et al. (2013). Aphotic N_2 fixation in the eastern tropical South Pacific Ocean. *PLoS ONE*, 8(12), e81265. <https://doi.org/10.1371/journal.pone.0081265>
- Boyer, T. P., Antonov, J. I., Baranova, O. K., Coleman, C., Garcia, H. E., Grodsky, A., et al. (2013). NOAA Atlas NESDIS 72. World Ocean Database. doi: <https://doi.org/10.7289/V5NZ85MT>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:101093340>
- Breitbarth, E., Oschlies, A., & LaRoche, J. (2007). Physiological constraints on the global distribution of *Trichodesmium* - effects of temperature on diazotrophy. *Biogeosciences*, 4(1), 53–61. <https://doi.org/10.5194/bg-4-53-2007>
- Breitbarth, E., Wohlers, J., Kläs, J., LaRoche, J., & Peeken, I. (2008). Nitrogen fixation and growth rates of *Trichodesmium* IMS-101 as a function of light intensity. *Marine Ecology Progress Series*, 359, 25–36. <https://doi.org/10.3354/meps07241>
- Capone, D. G., Burns, J. A., Montoya, J. P., Subramaniam, A., Mahaffey, C., Gunderson, T., et al. (2005). Nitrogen fixation by *Trichodesmium* spp.: An important source of new nitrogen to the tropical and subtropical North Atlantic Ocean. *Global Biogeochemical Cycles*, 19, GB2024. <https://doi.org/10.1029/2004GB002331>
- Capone, D. G., Zehr, J. P., Paerl, H. W., Bergman, B., & Carpenter, E. J. (1997). *Trichodesmium*, a globally significant marine cyanobacterium. *Science*, 276(5316), 1221–1229. <https://doi.org/10.1126/science.276.5316.1221>
- Church, M. J., Mahaffey, C., Letelier, R. M., Lukas, R., Zehr, J. P., & Karl, D. M. (2009). Physical forcing of nitrogen fixation and diazotroph community structure in the North Pacific subtropical gyre. *Global Biogeochemical Cycles*, 23, GB2020. <https://doi.org/10.1029/2008GB003418>
- Crombet, Y., Leblanc, K., Quéguiner, B., Moutin, T., Rimmel, P., Ras, J., et al. (2011). Deep silicon maxima in the stratified oligotrophic Mediterranean Sea. *Biogeosciences*, 8(2), 459–475. <https://doi.org/10.5194/bg-8-459-2011>
- de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., & Iudicone, D. (2004). Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *Journal of Geophysical Research*, 109, C12003. <https://doi.org/10.1029/2004JC002378>
- Delmont, T. O., Quince, C., Shaiber, A., Esen, Ö. C., Lee, S. T. M., Rappé, M. S., et al. (2018). Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nature Microbiology*, 3(7), 804–813. <https://doi.org/10.1038/s41564-018-0176-9>
- Dutheil, C., Aumont, O., Gorguès, T., Lorrain, A., Bonnet, S., Rodier, M., et al. (2018). Modelling N_2 fixation related to *Trichodesmium* sp.: Driving processes and impacts on primary production in the tropical Pacific Ocean. *Biogeosciences*, 15(14), 4333–4352. <https://doi.org/10.5194/bg-15-4333-2018>
- Eilers, P. H., & Goeman, J. J. (2004). Enhancing scatterplots with smoothed densities. *Bioinformatics*, 20(5), 623–628. <https://doi.org/10.1093/bioinformatics/btg454>

- Ellwood, M. J., Bowie, A. R., Baker, A., Gault-Ringold, M., Hassler, C., Law, C. S., et al. (2018). Insights into the biogeochemical cycling of iron, nitrate, and phosphate across a 5,300 km South Pacific zonal section (153°E–150°W). *Global Biogeochemical Cycles*, 32, 187–207. <https://doi.org/10.1002/2017GB005736>
- Eppley, R. W. (1972). Temperature and phytoplankton growth in the sea. *Fishery Bulletin*, 70(4), 1063–1085.
- Falkowski, P. G., Fenchel, T., & DeLong, E. F. (2008). The microbial engines that drive Earth's biogeochemical cycles. *Science*, 320(5879), 1034–1039. <https://doi.org/10.1126/science.1153213>
- Farnelid, H., Turk-Kubo, K., Muñoz-Marín, M. C., & Zehr, J. P. (2016). New insights into the ecology of the globally significant uncultured nitrogen-fixing symbiont UCYN-A. *Aquatic Microbial Ecology*, 77(3), 125–138. <https://doi.org/10.3354/ame01794>
- Fennel, K., Spitz, Y. H., Letelier, R. M., Abbott, M. R., & Karl, D. M. (2001). A deterministic model for N₂ fixation at stn. ALOHA in the subtropical North Pacific Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, 49(1–3), 149–174. [https://doi.org/10.1016/S0967-0645\(01\)00098-4](https://doi.org/10.1016/S0967-0645(01)00098-4)
- Follett, C. L., Dutkiewicz, S., Karl, D. M., Inomura, K., & Follows, M. J. (2018). Seasonal resource conditions favor a summertime increase in North Pacific diatom–diazotroph associations. *The ISME Journal*, 12(6), 1543–1557. <https://doi.org/10.1038/s41396-017-0012-x>
- Foster, R., Subramaniam, A., Mahaffey, C., Carpenter, E., Capone, D., & Zehr, J. (2007). Influence of the Amazon River plume on distributions of free-living and symbiotic cyanobacteria in the western tropical North Atlantic Ocean. *Limnology and Oceanography*, 52(2), 517–532. <https://doi.org/10.4319/lo.2007.52.2.0517>
- Fu, F.-X., Yu, E., Garcia, N. S., Gale, J., Luo, Y., Webb, E. A., & Hutchins, D. A. (2014). Differing responses of marine N₂ fixers to warming and consequences for future diazotroph community structure. *Aquatic Microbial Ecology*, 72(1), 33–46. <https://doi.org/10.3354/ame01683>
- Fu, F. X., Zhang, Y., Bell, P. R., & Hutchins, D. A. (2005). Phosphate uptake and growth kinetics of *Trichodesmium* (cyanobacteria) isolates from the North Atlantic Ocean and the Great Barrier Reef, Australia. *Journal of Phycology*, 41(1), 62–73. <https://doi.org/10.1111/j.1529-8817.2005.04063.x>
- Garcia, N. S., Fu, F. X., & Hutchins, D. A. (2013). Colimitation of the unicellular photosynthetic diazotroph *Crocosphaera watsonii* by phosphorus, light, and carbon dioxide. *Limnology and Oceanography*, 58(4), 1501–1512. <https://doi.org/10.4319/lo.2013.58.4.1501>
- Gregor, L., Kok, S., & Monteiro, P. (2017). Empirical methods for the estimation of Southern Ocean CO₂: Support vector and random forest regression. *Biogeosciences*, 14(23), 5551–5569. <https://doi.org/10.5194/bg-14-5551-2017>
- Halm, H., Lam, P., Ferdelman, T. G., Lavik, G., Dittmar, T., LaRoche, J., et al. (2012). Heterotrophic organisms dominate nitrogen fixation in the South Pacific Gyre. *The ISME Journal*, 6(6), 1238–1249. <https://doi.org/10.1038/ismej.2011.182>
- Harding, K., Turk-Kubo, K. A., Sipler, R. E., Mills, M. M., Bronk, D. A., & Zehr, J. P. (2018). Symbiotic unicellular cyanobacteria fix nitrogen in the Arctic Ocean. *Proceedings of the National Academy of Sciences*, 115(52), 13,371–13,375. <https://doi.org/10.1073/pnas.1813658115>
- Hashimoto, R., Watai, H., Miyahara, K., Sako, Y., & Yoshida, T. (2016). Spatial and temporal variability of unicellular diazotrophic cyanobacteria in the eastern Seto Inland Sea. *Fisheries Science*, 82(3), 459–471. <https://doi.org/10.1007/s12562-016-0983-y>
- Henke, B. A., Turk-Kubo, K. A., Bonnet, S., & Zehr, J. P. (2018). Distributions and Abundances of sublineages of the N₂-Fixing cyanobacterium Candidatus Atelocyanobacterium thalassa (UCYN-A) in the new caledonian coral lagoon. *Frontiers in Microbiology*, 9, 554. <https://doi.org/10.3389/fmicb.2018.00554>
- Hoffman, B. M., Lukoyanov, D., Yang, Z. Y., Dean, D. R., & Seefeldt, L. C. (2014). Mechanism of nitrogen fixation by nitrogenase: The next stage. *Chemical Reviews*, 114(8), 4041–4062. <https://doi.org/10.1021/cr400641x>
- Holl, C. M., Waite, A. M., Pesant, S., Thompson, P. A., & Montoya, J. P. (2007). Unicellular diazotrophy as a source of nitrogen to Leeuwin Current coastal eddies. *Deep Sea Research Part II: Topical Studies in Oceanography*, 54(8), 1045–1054. <https://doi.org/10.1016/j.dsr2.2007.02.002>
- Hood, R. R., Bates, N. R., Capone, D. G., & Olson, D. B. (2001). Modeling the effect of nitrogen fixation on carbon and nitrogen fluxes at BATS. *Deep Sea Research Part II: Topical Studies in Oceanography*, 48(8), 1609–1648. [https://doi.org/10.1016/S0967-0645\(00\)00160-0](https://doi.org/10.1016/S0967-0645(00)00160-0)
- Hood, R. R., Coles, V. J., & Capone, D. (2004). Modeling the distribution of *Trichodesmium* and nitrogen fixation in the Atlantic Ocean. *Journal of Geophysical Research*, 109, C06006. <https://doi.org/10.1029/2002JC001753>
- Jacq, V., Ridame, C., L'Helguen, S., Kaczmar, F., & Saliot, A. (2014). Response of the unicellular diazotrophic cyanobacterium *Crocosphaera watsonii* to iron limitation. *PLoS ONE*, 9(1), e86749. <https://doi.org/10.1371/journal.pone.0086749>
- Jyothibabu, R., Karnan, C., Jagadeesan, L., Arunpandi, N., Pandiarajan, R., Muraleedharan, K., & Balachandran, K. (2017). *Trichodesmium* blooms and warm-core ocean surface features in the Arabian Sea and the Bay of Bengal. *Marine Pollution Bulletin*, 121(1–2), 201–215. <https://doi.org/10.1016/j.marpolbul.2017.06.002>
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77(3), 437–471. [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2)
- Karl, D., Letelier, R., Tupas, L., Dore, J., Christian, J., & Hebel, D. (1997). The role of nitrogen fixation in biogeochemical cycling in the subtropical North Pacific Ocean. *Nature*, 388(6642), 533–538. <https://doi.org/10.1038/41474>
- Karl, D., Michaels, A., Bergman, B., Capone, D., Carpenter, E., Letelier, R., et al. (2002). Dinitrogen fixation in the world's oceans. *Biogeochemistry*, 57(1), 47–98. <https://doi.org/10.1023/A:1015798105851>
- Knapp, A. N. (2012). The sensitivity of marine N₂ fixation to dissolved inorganic nitrogen. *Frontiers in Microbiology*, 3. <https://doi.org/10.3389/fmicb.2012.00374>
- Knapp, A. N., Casciotti, K. L., Berelson, W. M., Prokopenko, M. G., & Capone, D. G. (2016). Low rates of nitrogen fixation in eastern tropical South Pacific surface waters. *Proceedings of the National Academy of Sciences*, 113(16), 4398–4403. <https://doi.org/10.1073/pnas.1515641113>
- Krupke, A., Musat, N., LaRoche, J., Mohr, W., Fuchs, B. M., Amann, R. I., et al. (2013). In situ identification and N₂ and C fixation rates of uncultivated cyanobacteria populations. *Systematic and Applied Microbiology*, 36(4), 259–271. <https://doi.org/10.1016/j.syapm.2013.02.002>
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3), 18–22.
- Luo, Y. W., Doney, S. C., Anderson, L. A., Benavides, M., Berman-Frank, I., Bode, A., et al. (2012). Database of diazotrophs in global ocean: Abundance, biomass and nitrogen fixation rates. *Earth System Science Data*, 4(1), 47–73. <https://doi.org/10.5194/essd-4-47-2012>
- Luo, Y. W., Lima, I. D., Karl, D. M., Deutsch, C. A., & Doney, S. C. (2014). Data-based assessment of environmental controls on global marine nitrogen fixation. *Biogeosciences*, 11(3), 691–708. <https://doi.org/10.5194/bg-11-691-2014>
- Martínez-Pérez, C., Mohr, W., Löscher, C. R., Dekaezemacker, J., Littmann, S., Yilmaz, P., et al. (2016). The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle. *Nature Microbiology*, 1(11), 16,163. <https://doi.org/10.1038/nmicrobiol.2016.163>

- Mather, R. L., Reynolds, S. E., Wolff, G. A., Williams, R. G., Torres-Valdes, S., Woodward, E. M. S., et al. (2008). Phosphorus cycling in the North and South Atlantic Ocean subtropical gyres. *Nature Geoscience*, 1(7), 439–443. <https://doi.org/10.1038/ngeo232>
- Moisander, P. H., Beinart, R. A., Hewson, I., White, A. E., Johnson, K. S., Carlson, C. A., et al. (2010). Unicellular cyanobacterial distributions broaden the oceanic N₂ fixation domain. *Science*, 327(5972), 1512–1514. <https://doi.org/10.1126/science.1185468>
- Monteiro, F. M., Follows, M. J., & Dutkiewicz, S. (2010). Distribution of diverse nitrogen fixers in the global ocean. *Global Biogeochemical Cycles*, 24, GB3017. <https://doi.org/10.1029/2009GB003731>
- Moore, C. M., Mills, M. M., Achterberg, E. P., Geider, R. J., LaRoche, J., Lucas, M. I., et al. (2009). Large-scale distribution of Atlantic nitrogen fixation controlled by iron availability. *Nature Geoscience*, 2(12), 867–871. <https://doi.org/10.1038/ngeo667>
- Moore, J. K., Lindsay, K., Doney, S. C., Long, M. C., & Misumi, K. (2013). Marine ecosystem dynamics and biogeochemical cycling in the Community Earth System Model [CESM1 (BGC)]: Comparison of the 1990s with the 2090s under the RCP4. 5 and RCP8. 5 scenarios. *Journal of Climate*, 26(23), 9291–9312. <https://doi.org/10.1175/JCLI-D-12-00566.1>
- Moreira-Coello, V., Mouriño-Carballido, B., Marañón, E., Fernández-Carrera, A., Bode, A., & Varela, M. M. (2017). Biological N₂ Fixation in the Upwelling Region off NW Iberia: Magnitude, Relevance, and Players. *Frontiers in Marine Science*, 4, 303. <https://doi.org/10.3389/fmars.2017.00303>
- Mulholland, M. R., Bernhardt, P. W., Widner, B. N., Selden, C. R., Chappell, P. D., Clayton, S., et al. (2019). High Rates of N₂ Fixation in Temperate, Western North Atlantic Coastal Waters Expand the Realm of Marine Diazotrophy. *Global Biogeochemical Cycles*, 33, 826–840. <https://doi.org/10.1029/2018GB006130>
- Paulsen, H., Ilyina, T., Six, K. D., & Stemmler, I. (2017). Incorporating a prognostic representation of marine nitrogen fixers into the global ocean biogeochemical model HAMOC. *Journal of Advances in Modeling Earth Systems*, 9, 438–464. <https://doi.org/10.1002/2016MS000737>
- Poulton, A. J., Stinchcombe, M. C., & Quartly, G. D. (2009). High numbers of *Trichodesmium* and diazotrophic diatoms in the southwest Indian Ocean. *Geophysical Research Letters*, 36, L15610. <https://doi.org/10.1029/2009GL039719>
- Pyle, A. E. (2011). *Light dependant growth and nitrogen fixation rates in the Hemialus haukii and Hemialus membranaceus diatom-diazotroph associations*, (Master Thesis). Retrieved from Texas ScholarWorks. Austin, USA: University of Texas. <http://hdl.handle.net/2152/ETD-UT-2011-12-4629>
- Sañudo-Wilhelmy, S. A., Kustka, A. B., Gobler, C. J., Hutchins, D. A., Yang, M., Lwiza, K., et al. (2001). Phosphorus limitation of nitrogen fixation by *Trichodesmium* in the central Atlantic Ocean. *Nature*, 411(6833), 66–69. <https://doi.org/10.1038/35075041>
- Sargent, E. C., Hitchcock, A., Johansson, S. A., Langlois, R., Moore, C. M., LaRoche, J., et al. (2016). Evidence for polyploidy in the globally important diazotroph *Trichodesmium*. *FEMS Microbiology Letters*, 363(21). <https://doi.org/10.1093/femsle/fnw244>
- Shiozaki, T., Bombar, D., Riemann, L., Hashihama, F., Takeda, S., Yamaguchi, T., et al. (2017). Basin scale variability of active diazotrophs and nitrogen fixation in the North Pacific, from the tropics to the subarctic Bering Sea. *Global Biogeochemical Cycles*, 31, 996–1009. <https://doi.org/10.1002/2017GB005681>
- Shiozaki, T., Bombar, D., Riemann, L., Sato, M., Hashihama, F., Kodama, T., et al. (2018). Linkage between dinitrogen fixation and primary production in the oligotrophic South Pacific Ocean. *Global Biogeochemical Cycles*, 32, 1028–1044. <https://doi.org/10.1029/2017GB005869>
- Shiozaki, T., Chen, Y. L. L., Lin, Y. H., Taniuchi, Y., Sheu, D. S., Furuya, K., & Chen, H. Y. (2014). Seasonal variations of unicellular diazotroph groups A and B, and *Trichodesmium* in the northern South China Sea and neighboring upstream Kuroshio Current. *Continental Shelf Research*, 80, 20–31. <https://doi.org/10.1016/j.csr.2014.02.015>
- Shiozaki, T., Fujiwara, A., Ijichi, M., Harada, N., Nishino, S., Nishi, S., et al. (2018). Diazotroph community structure and the role of nitrogen fixation in the nitrogen cycle in the Chukchi Sea (western Arctic Ocean). *Limnology and Oceanography*, 63(5), 2191–2205. <https://doi.org/10.1002/lno.10933>
- Shiozaki, T., Ijichi, M., Kodama, T., Takeda, S., & Furuya, K. (2014). Heterotrophic bacteria as major nitrogen fixers in the euphotic zone of the Indian Ocean. *Global Biogeochemical Cycles*, 28, 1096–1110. <https://doi.org/10.1002/2014GB004886>
- Shiozaki, T., Kodama, T., & Furuya, K. (2014). Large-scale impact of the island mass effect through nitrogen fixation in the western South Pacific Ocean. *Geophysical Research Letters*, 41, 2907–2913. <https://doi.org/10.1002/2014GL059835>
- Shiozaki, T., Kondo, Y., Yuasa, D., & Takeda, S. (2018). Distribution of major diazotrophs in the surface water of the Kuroshio from northeastern Taiwan to south of mainland Japan. *Journal of Plankton Research*, 40(4), 407–419. <https://doi.org/10.1093/plankt/fby027>
- Shiozaki, T., Nagata, T., Ijichi, M., & Furuya, K. (2015). Nitrogen fixation and the diazotroph community in the temperate coastal region of the northwestern North Pacific. *Biogeosciences*, 12(15), 4751–4764. <https://doi.org/10.5194/bg-12-4751-2015>
- Sohm, J. A., Mahaffey, C., & Capone, D. G. (2008). Assessment of relative phosphorus limitation of *Trichodesmium* spp. in the North Pacific, North Atlantic, and the north coast of Australia. *Limnology and Oceanography*, 53(6), 2495–2502. <https://doi.org/10.4319/lo.2008.53.6.2495>
- Sohm, J. A., Webb, E. A., & Capone, D. G. (2011). Emerging patterns of marine nitrogen fixation. *Nature Reviews Microbiology*, 9(7), 499–508. <https://doi.org/10.1038/nrmicro2594>
- Srokosz, M. A., & Quartly, G. D. (2013). The Madagascar bloom: A serendipitous study. *Journal of Geophysical Research: Oceans*, 118, 14–25. <https://doi.org/10.1029/2012JC008339>
- Stenegren, M., Caputo, A., Berg, C., Bonnet, S., & Foster, R. A. (2018). Distribution and drivers of symbiotic and free-living diazotrophic cyanobacteria in the western tropical South Pacific. *Biogeosciences*, 15(5), 1559–1578. <https://doi.org/10.5194/bg-15-1559-2018>
- Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., et al. (2015). Structure and function of the global ocean microbiome. *Science*, 348(6237). <https://doi.org/10.1126/science.1261359>
- Tang, W., Li, Z., & Cassar, N. (2019). Machine learning estimates of global oceanic nitrogen fixation. *Journal of Geophysical Research: Biogeosciences*, 124, 717–730. <https://doi.org/10.1029/2018JG004828>
- Tang, W., Wang, S., Fonseca-Batista, D., Dehairs, F., Gifford, S., Gonzalez, A. G., et al. (2019). Revisiting the distribution of oceanic N₂ fixation and estimating diazotrophic contribution to marine production. *Nature Communications*, 10(1), 831. <https://doi.org/10.1038/s41467-019-08640-0>
- Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93(4), 485–498. <https://doi.org/10.1175/BAMS-D-11-00094.1>
- Thompson, A., Carter, B. J., Turk-Kubo, K., Malfatti, F., Azam, F., & Zehr, J. P. (2014). Genetic diversity of the unicellular nitrogen-fixing cyanobacteria UCYN-A and its prymnesiophyte host. *Environmental Microbiology*, 16(10), 3238–3249. <https://doi.org/10.1111/1462-2920.12490>
- Turk-Kubo, K. A., Connell, P. E., Caron, D. A., Hogan, M. E., Farnelid, H. M., & Zehr, J. P. (2018). *In situ* diazotroph population dynamics under different resource ratios in the North Pacific Subtropical Gyre. *Frontiers in Microbiology*, 9, 1616. <https://doi.org/10.3389/fmicb.2018.01616>

- Villareal, T. A. (1991). Nitrogen-fixation by the cyanobacterial symbiont of the diatom genus *Hemiaulus*. *Marine Ecology Progress Series*, 76(2), 201–204. <https://doi.org/10.3354/meps076201>
- Villareal, T. A., Brown, C. G., Brzezinski, M. A., Krause, J. W., & Wilson, C. (2012). Summer diatom blooms in the North Pacific subtropical gyre: 2008–2009. *PLoS ONE*, 7(4), e33109. <https://doi.org/10.1371/journal.pone.0033109>
- Wang, W. L., Moore, J. K., Martiny, A. C., & Primeau, F. W. (2019). Convergent estimates of marine nitrogen fixation. *Nature*, 566(7743), 205–211. <https://doi.org/10.5281/zenodo.2020377>
- Ward, B. A., Dutkiewicz, S., Moore, C. M., & Follows, M. J. (2013). Iron, phosphorus, and nitrogen supply ratios define the biogeography of nitrogen fixation. *Limnology and Oceanography*, 58(6), 2059–2075. <https://doi.org/10.4319/lo.2013.58.6.2059>
- Weber, S. C., Carpenter, E. J., Coles, V. J., Yager, P. L., Goes, J., & Montoya, J. P. (2016). Amazon River influence on nitrogen fixation and export production in the western tropical North Atlantic. *Limnology and Oceanography*, 62(2), 618–631. <https://doi.org/10.1002/lno.10448>
- Wen, Z., Lin, W., Shen, R., Hong, H., Kao, S. J., & Shi, D. (2017). Nitrogen fixation in two coastal upwelling regions of the Taiwan Strait. *Scientific Reports*, 7(1), 17,601. <https://doi.org/10.3389/fmars.2017.00303>
- Westberry, T. K., & Siegel, D. A. (2006). Spatial and temporal distribution of *Trichodesmium* blooms in the world's oceans. *Global Biogeochemical Cycles*, 20, GB4016. <https://doi.org/10.1029/2005GB002673>
- White, A. E., Watkins-Brandt, K. S., & Church, M. J. (2018). Temporal Variability of *Trichodesmium* spp. and Diatom-Diazotroph Assemblages in the North Pacific Subtropical Gyre. *Frontiers in Marine Science*, 5. <https://doi.org/10.3389/fmars.2018.00027>
- Worden, A. Z., Follows, M. J., Giovannoni, S. J., Wilken, S., Zimmerman, A. E., & Keeling, P. J. (2015). Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science*, 347(6223). <https://doi.org/10.1126/science.1257594>
- Wu, C., Kan, J., Liu, H., Pujari, L., Guo, C., Wang, X., & Sun, J. (2019). Heterotrophic Bacteria Dominate the Diazotrophic Community in the Eastern Indian Ocean (EIO) during Pre-Southwest Monsoon. *Microbial Ecology*, 1–16. <https://doi.org/10.1007/s00248-019-01355-1>
- Wu, J., Sunda, W., Boyle, E. A., & Karl, D. M. (2000). Phosphate depletion in the western North Atlantic Ocean. *Science*, 289(5480), 759–762. <https://doi.org/10.1126/science.289.5480.759>
- Zehr, J. P. (2011). Nitrogen fixation by marine cyanobacteria. *Trends in Microbiology*, 19(4), 162–173. <https://doi.org/10.1016/j.tim.2010.12.004>
- Zehr, J. P., Mellon, M. T., & Zani, S. (1998). New nitrogen-fixing microorganisms detected in oligotrophic oceans by amplification of nitrogenase (*nifH*) genes. *Applied and Environmental Microbiology*, 64(9), 3444–3450.