Data-Driven Modeling of the Distribution of Diazotrophs in the Global Ocean
Weiyi Tang, Nicolas Cassar

To cite this version:

HAL Id: hal-02920511
https://hal.science/hal-02920511
Submitted on 11 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
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:

Received 1 JUL 2019
Accepted 29 OCT 2019
Accepted article online 9 NOV 2019
Published online 13 NOV 2019

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 ecophysiology 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.,
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; Krapke 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, Fujikawa 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 (i.e., *Crocosphaera*), and *Richelia*. We conclude by comparing the geographical distributions of our data-derived estimates to trait-based model simulations.

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
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 \textit{Trichodesmium}, UCYN-A, UCYN-B, and \textit{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 \textit{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 1. Global distributions of the depth-integrated \textit{nifH} gene abundances of four major diazotrophs (a) \textit{Trichodesmium}, (b) UCYN-A, (c) UCYN-B, and (d) \textit{Richelia}.
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 N2 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 ~26°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 monthly abundance of the four diazotroph groups in the global ocean with a 2°×2° resolution. A thorough description of the RF method and its applications can be found in Breiman, 2001.
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 ~30°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 (Sahudo-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; Martinez-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* (Duthieil 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:
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 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.
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 ecophysologies 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

<table>
<thead>
<tr>
<th>Species</th>
<th>T ($^\circ$C)</th>
<th>Light</th>
<th>$K_s$ (P)</th>
<th>$K_s$ (Fe)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Trichodesmium</em></td>
<td>20-34 (~27)*</td>
<td>180*</td>
<td>0.14-0.22*</td>
<td>0.1-0.3*</td>
</tr>
<tr>
<td>UCYN-A</td>
<td>18-32 (~26)*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UCYN-B</td>
<td>24-32 (~30)*</td>
<td>76*</td>
<td>0.1-0.25*</td>
<td>6.95*</td>
</tr>
<tr>
<td>DDA</td>
<td>N/A</td>
<td>100*</td>
<td>0.13*</td>
<td>0.5*</td>
</tr>
</tbody>
</table>

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 Takuehi 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 CMIPS, NASA Ocean Color, NCEP/NCAR reanalysis, World Ocean Atlas, and Ifremer for processing and distributing their products. The updated database 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 French government (ANR‐19) and cofunded by a CAREER grant (1350710). N. C. is also supported by the French government (ANR‐19) and cofunded by a CAREER grant (1350710). W. T. and N. C. are supported by the French government (ANR‐19) and cofunded by a CAREER grant (1350710). N. C. is also supported by the French government (ANR‐19) and cofunded by a CAREER grant (1350710). W. T. and N. C. are supported by the French government (ANR‐19) and cofunded by a CAREER grant (1350710). N. C. is also supported by the French government (ANR‐19) and cofunded by a CAREER grant (1350710).

References


