Data-driven modeling of the distribution of diazotrophs in the global ocean

**2019 OCB summer workshop
24-27, 2019 in Woods Hole**

Weiyi Tang[1] & Nicolas Cassar[1,2]

1. Earth and Ocean Sciences, Nicholas School of the Environment, Duke University, Durham, NC
2. Laboratoire des Sciences de l'Environnement Marin (LEMAR), Institut Universitaire Européen de la Mer (IUEM), Brest, France
Contact: **weiyi.tang@duke.edu**

Introduction

- As key microorganisms in the ocean, diazotrophs convert N_2 gas into bioavailable nitrogen (N_2 fixation), thereby relieving nitrogen limitation and supporting marine production in many regions of the world's oceans (1-2). Despite the central role of diazotrophs, factors controlling on their distributions remain elusive (3).

- Luo et al (2012) compiled the first database of diazotrophs in the global ocean (4). The number of observations has rapidly expanded since then, with new data collected in coastal, aphotic, and polar waters (5-7). In light of such progress, updating the previous database and re-evaluating the factors controlling on their distributions represents a timely effort.

- Prognostic parameterizations have been used to model the regional and global distribution of diazotrophs (8-9). However, most models focus on *Trichodesmium* despite the increasing appreciation of diazotrophs' diversity.

- Machine learning techniques have increasingly been applied to marine sciences, e.g. estimating global N_2 fixation rates (10) and net community production (11).

- **Aim: To identify strong predictors of diazotroph abundances, to construct data-driven global biogeographies of various diazotrophs using a machine learning method and finally to compare our estimates to the ones derived by trait-based models.**

Methods and data

1. Global diazotrophs dataset is updated (Figure 1), more than doubling the size of observations compared to Luo et al (2012) dataset.
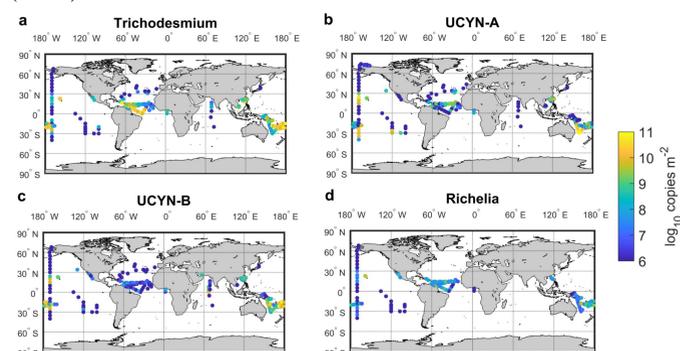

Figure 1. Global distributions of four major diazotrophs' *nifH* gene copies quantified using qPCR assays (a) *Trichodesmium*, (b) UCYN-A, (c) UCYN-B and (d) *Richelia*.

2. **Correlation analysis (Figure 2 and Figure 3).**
The volumetric abundances of diazotrophs are compared with field-measured environmental variables hypothesized to regulate diazotrophy, such as depth, temperature and nutrients.

3. **Depth-integrated diazotroph abundances are matched to various environmental factors spatiotemporally.**
Daily: solar radiation; wind speed (NCEP/NCAR)
8-day: sea surface temperature, photosynthetically available radiation; chlorophyll_*a* concentration (NASA Ocean Color)
Monthly climatology: sea surface salinity; nutrients; oxygen concentration (WOA); mixed layer depth (12)
Annual climatology: surface iron concentration (13)
Data are binned into $2° \times 2°$ resolution after matching.

4. **Random forest (RF) is applied to estimate depth-integrated abundances of four major diazotrophs using compiled environmental factors (Figure 5).**
Construct 100 additional datasets using a bootstrap method. Models are built with randomly selected training data (70% of each dataset) and evaluated using the test data (30% of each dataset). The model ensemble mean is presented.
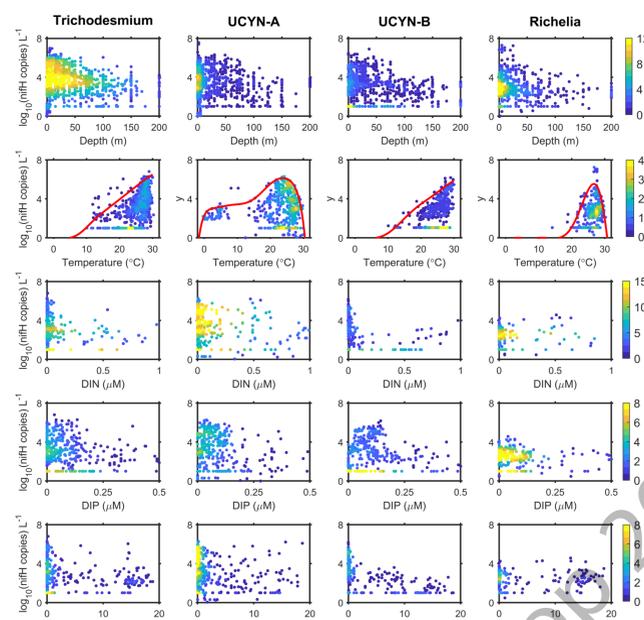
Results

Figure 2. Volumetric diazotroph abundances vs contemporaneously field-observed environmental properties. Points are color coded for density of observations (14).
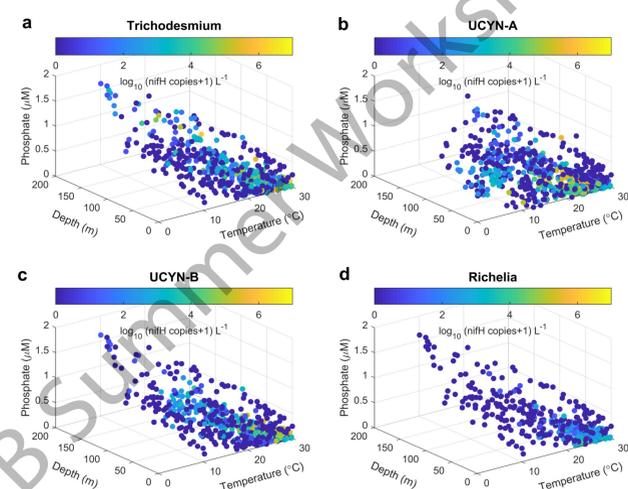

Figure 3. Controls on diazotroph abundances by multiple environmental properties.

Distinct environmental controls on four major diazotrophs:
1. **No clear depth separation is found for diazotrophs.**
2. **Temperature sets upper bounds on the diazotroph abundances.**
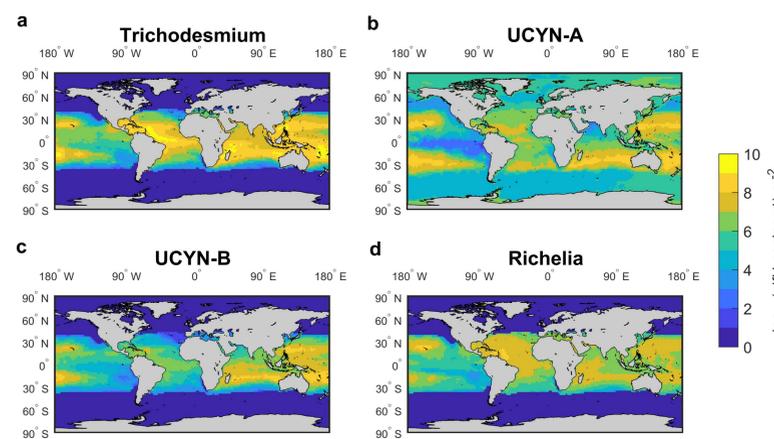3. **Light and nutrients modulate diazotroph abundances globally.**

Figure 4. RF-predicted global distributions of (a) *Trichodesmium*, (b) UCYN-A, (c) UCYN-B and (d) *Richelia*.

1. **Machine learning estimates agree well with observed diazotroph abundances independent of the training dataset.**
2. **Diazotrophs show high abundance in the western subtropical Pacific. In contrast, UCYN-A is the only one predicted in cold polar waters.**
3. **Hotspots of diazotroph predicted in the southern Indian ocean and South Atlantic warrant future field investigations.**
4. **Large discrepancies exist among various models notably in the eastern tropical Pacific, temperate and polar waters.**
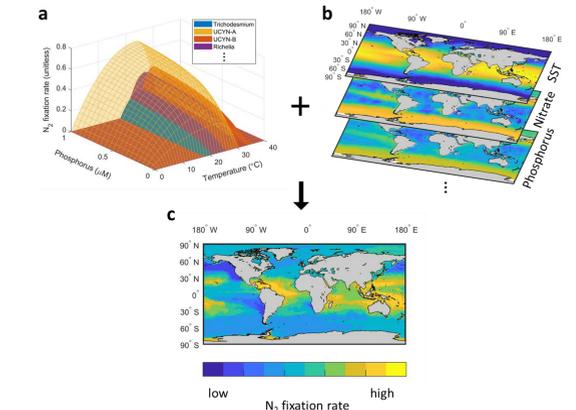
Results

Figure 5. **Conceptual diagram of an alternative approach to model marine N_2 fixation accounting for the granularity in the ecophysiologies of diazotrophs.** (a) Distinct environmental controls among diazotrophs are coupled with (b) the distributions of environmental properties to simulate (c) global marine N_2 fixation.

Caveats

- Due to the limited number of qPCR observations, we restricted our study to four major diazotrophs. For example, non-cyanobacterial diazotrophs are not included.

- *nifH* gene copy numbers are not necessarily equal to cell numbers. Unclear connections from the presence of *nifH* gene to *nifH* gene expression and N_2 fixation activity.

- Mismatch of diazotroph abundances with predictors in space and time: e.g. climatologies were used if contemporaneous predictors were not available.

Conclusions

- The distribution of diazotrophs is controlled by a complex interplay of factors with diazotroph groups displaying distinct biogeographic niches.

- Machine learning estimates identify regions worthy of further investigation because they may boast diazotrophic hotspots, remain presently undersampled, or produce large discrepancies in model simulations.

- Combining diazotrophs' individual distributions and their activities may be a potential strategy to represent global distribution of N_2 fixation.

References

1. Sohm, J. A., Webb, E. A., & Capone, D. G. (2011). Emerging patterns of marine nitrogen fixation. *Nature Reviews Microbiology*, 9(7), 499-508.
2. Karl, D., Michaels, A., Bergman, B., Capone, D., Carpenter, E., Letelier, R., ... & Stal, L. (2002). Dinitrogen fixation in the world's oceans. *Biogeochemistry*, 57(1), 47-98.
3. Gruber, N. (2016). Elusive marine nitrogen fixation. *PNAS*, 113(16), 4246-4248.
4. Luo, Y. W., Doney, S. C., Anderson, L. A., Benavides, M., Berman-Frank, I., Bode, A., ... & Mulholland, M. R. (2012). Database of Diazotrophs in Global Ocean: Abundance, Biomass, and Nitrogen Fixation Rates. *Earth System Science Data*, 4(1), 47-73.
5. Tang, W., Wang, S., Fonseca-Batista, D., Dehairs, F., Gifford, S., Gonzalez, A. G., ... & Cassar, N. (2019). Revisiting the distribution of oceanic N_2 fixation and estimating diazotrophic contribution to marine production. *Nature Communications*, 10(1), 831.
6. Fernandez, C., Farías, L., & Ulloa, O. (2011). Nitrogen fixation in denitrified marine waters. *PloS one*, 6(6), e20539.
7. 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. *PNAS*, 115(52), 13371-13375.
8. Hood, R. R., Coles, V. J., & Capone, D. (2004). Modeling the distribution of *Trichodesmium* and nitrogen fixation in the Atlantic Ocean. *JGR*, 109(C6).
9. Paulsen, H., Ilyina, T., Six, K. D., & Stemmler, I. (2017). Incorporating a prognostic representation of marine nitrogen fixers into the global ocean biogeochemical model HAMOCC. *Journal of Advances in Modeling Earth Systems*, 9(1), 438-464.
10. Tang, W., Li, Z., & Cassar, N. (2019). Machine learning estimates of global marine nitrogen fixation. *JGR: Biogeosciences*, 124, 717– 730.
11. Li, Z., & Cassar, N. (2016). Satellite estimates of net community production based on O_2/Ar observations and comparison to other estimates. *Global Biogeochemical Cycles*, 30(5), 735-752.
12. 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. *JGR: Oceans*, 109(C12).
13. 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.
14. Eilers, P. H., & Goeman, J. J. (2004). Enhancing scatterplots with smoothed densities. *Bioinformatics*, 20(5), 623-628.

Acknowledgement