Challenges and opportunities in connecting gene count observations with ocean biogeochemical models: Reply to Zehr and Riemann (2023)

Author(s):
Meiler, Simona; Britten, Gregory L.; Dutkiewicz, Stephanie; Moisander, Pia H.; Follows, Michael J.

Publication date:
2023-06

Permanent link:
https://doi.org/10.3929/ethz-b-000612880

Originally published in:
Limnology and Oceanography 68(6), https://doi.org/10.1002/lno.12363

LIMNOLOGY and OCEANOGRAPHY

ASLO

Limnol. Oceanogr. 68, 2023, 1413–1416

doi: 10.1002/lno.12363

Challenges and opportunities in connecting gene count observations with ocean biogeochemical models: Reply to Zehr and Riemann (2023)

Simona Meiler,[1*] **Gregory L. Britten**,[2,3] **Stephanie Dutkiewicz**,[2] **Pia H. Moisander**,[4] **Michael J. Follows**[2]

[1]Department of Environmental Systems Science, ETH Zurich, Zurich, Switzerland
[2]Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[3]Department of Biology, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, USA
[4]Department of Biology, University of Massachusetts Dartmouth, North Dartmouth, Massachusetts, USA

As authors of Meiler et al. (2022), we welcome Zehr and Riemann's (2023) comment and discussion. We agree, of course, with the general statement that "quantification of gene copy numbers is valuable in marine microbial ecology" and wish to clarify that one of the purposes of Meiler et al. (2022) was to address the *specific* challenge of using a compilation of quantitative polymerase chain reaction (qPCR) *nifH* data to evaluate the skill of biogeochemical models. In that particular case, the data were most helpful in constraining the range of diazotrophs, but several sources of uncertainty limited more detailed quantitative evaluations. This was not intended to imply a lack of value or promise for such applications of qPCR data: we believe that testing and constraining biogeochemical and ecological models will be an important application of qPCR data, yet the quantitative interface between molecular data and biogeochemical models remains at its infancy. In the following, we first provide a background perspective for the Meiler et al. (2022) study, pointing out why observations and simulations are rooted in different currencies. We then discuss in more detail some of the specific points raised by Zehr and Riemann (2023) and highlight why further efforts toward intercalibration of currencies used to measure and simulate marine microbial populations is particularly significant if we are to fully exploit the data in biogeochemical and climate modeling applications. We end by summarizing some potentially fruitful avenues for future effort stimulated by this dialog.

Background: Observing and modeling diazotroph populations in the ocean

Nitrogen fixation is of global biogeochemical and ecological significance. Spatial and temporal variations in marine nitrogen fixation can significantly alter ecosystem structure and function (Karl et al. 1997). However, our collective knowledge of the distribution of diazotrophs over large spatial and temporal scales is still limited by the sparsity of observations. Molecular methods, including qPCR, provide an efficient and valuable way to increase empirical knowledge of diazotroph biogeography. In parallel, biogeochemical models provide a tool for synthesizing our understanding of the controls on diazotrophy by simulating the biochemical environment and community interactions within which diazotrophs exist. Such models are also employed to predict impacts of global change. Hence, it is important to test models and simulations with observational data.

Several current climate and carbon cycle models resolve an explicit, diazotrophic (nitrogen-fixing) class of plankton (Aumont and Bopp 2006; Dunne et al. 2013; Moore et al. 2013) with a few resolving diversity within diazotrophs (Monteiro et al. 2010; Dutkiewicz et al. 2012; Stukel et al. 2014; Coles et al. 2017; Follett et al. 2018). However, differences in the parameterization of nitrogen fixers in models used for climate and global change projections leads to extremely divergent predictions of diazotroph biogeography now and in the future (Landolfi et al. 2015; Wrightson and Tagliabue 2020). Several models predict nitrogen fixation mostly in upwelling regimes, while others predict that it largely occurs in downwelling subtropical gyres (*see* Wrightson and Tagliabue 2020). These predictions cannot all be correct. In parallel to numerical simulations, efforts have been made to interpret patterns of nitrogen fixation in the ocean based on established ecological theory (Dutkiewicz et al. 2012; Ward et al. 2013). A challenge is to test (not necessarily validate) these simulations and theories using the observations. Which models can be rejected, and which are potentially useful? Can

*Correspondence: simona.meiler@usys.ethz.ch

Author Contribution Statement: S.M., G.L.B., S.D., P.H.M., and M.J.F. contributed to the writing, reviewing and editing of the manuscript.

we use empirical data to evaluate optimal parameter values in plausible models?

Molecular metrics have great potential to provide global-scale surveys, benchmarking the current biogeography and monitoring changes in the populations of marine microbes (notably *Tara* Oceans; Pesant et al. 2015). They are efficient, increasingly cost-effective, and the most likely route for providing the data density and frequency required to test or constrain basin and global-scale models (Biller et al. 2018). However, the application of such data to testing and constraining models is still in its infancy. In Meiler et al. (2022), we explored whether we could quantitatively discriminate between simulated biogeographies using a published global qPCR data compilation (Tang and Cassar 2019). We found that a presence–absence metric was effective in evaluating predicted diazotroph ranges, but uncertainties in the intercalibration of gene and biomass units limited the ability to evaluate more subtle features of simulated biomass concentrations. One particular source of uncertainty is the intercalibration of the currencies used to quantify population densities: qPCR provides a gene density (e.g., *nifH* gene copies L^{-1}) while biogeochemical models are necessarily formulated in terms of biomass (e.g., moles C L^{-1}).

Why don't models and data use the same currency? This is due to the fundamental nature of the underlying methodologies. At the heart of the population models employed in biogeochemical simulations are statements of mass conservation: mass (e.g., atoms of carbon) cannot be created or destroyed, only reallocated between living biomass, detritus, and inorganic forms. This provides a powerful physical constraint that is essential to the success of such models making biomass the natural currency. In parallel, qPCR is clearly a gene-based technology and the enumeration of genes is its natural currency. Genes such as *nifH* form a minute and variable fraction of total biomass so to quantitatively interface gene densities and modeled biomass an intercalibration is required: either data must be re-expressed as biomass, or modeled biomass must be related to gene densities (Coles et al. 2017). Critically, as both Meiler et al. (2022) and Zehr and Riemann (2023) discuss, this intercalibration requires knowledge of several factors including gene copies per cell, and biomass per cell, which are not yet well constrained. The intercalibration issue is not unique to gene densities (as also noted by Zehr and Riemann 2023): The synthesis of published biogeography compilations of several phytoplankton functional types, relying mostly on microscopy, faces similar issues (Buitenhuis et al. 2013, and references therein) and uncertainty is also present in conversions between flow cytometry pico- and pico-nano-sized cell counts to biomass (Ribalet et al. 2019).

In the past, evaluations of simulated diazotroph biogeography (or nitrogen fixation) have typically been carried out qualitatively by eye-balled comparison with available observations. Meiler et al. (2022) sought to develop a more rigorous, quantitative test of modeled diazotroph biogeography using the Tang and Cassar (2019) qPCR *nifH* data set, along with ancillary published data which provided information on the conversion between biomass and gene density. Uncertainties in *nifH* genes-per-cell (influenced by polyploidy), as well as cell quotas, along with data aggregation, lead to wide ranges of uncertainty in the calibrations with implications for model testing. Zehr and Riemann (2023) argue that the study may lead readers to under-value *nifH* gene density data, highlighting several specific concerns. In the following section we address those points and identify avenues for positive future action.

Zehr and Riemann's comments: Challenges and opportunities

Zehr and Riemann (2023) suggest that previous work challenges the significant uncertainties found by Meiler et al. (2022), because (referring to earlier publications) spatial patterns of gene abundance "echo the distribution of N_2 fixation rates from compilations of observations, inverse models based on nutrient distributions, and ecosystem models." However, these are qualitative statements and the cited models and data sets are also subject to similar calibration uncertainties and sparser data constraints. For example, the "close relationship" referred to in Luo et al. (2012) is based on a qualitative interpretation of sparse data coverage and high spatial variability (their Figs. 8a, 9a). The maps of *nifH*-based biomass in that paper assume a *nifH* : cell ratio of 1 : 1 which does not consider the potential for polyploidy (Krupke et al. 2013; White et al. 2018; Gradoville et al. 2022), leading to very high uncertainty. We reiterate that one of the purposes of the Meiler et al. (2022) study was to seek avenues to go beyond such qualitative evaluations and bring more rigorous, quantitative approaches to model evaluations which explicitly take into account uncertainties. This is necessary if we are to achieve the goal of robust and valuable models for global change studies. Although several studies have addressed the importance of rigorous model-skill assessment (Doney et al. 2009; Friedrichs et al. 2007), they have not focused on the specific uncertainties involved in using gene-count observations.

Meiler et al. (2022) compiled published data on gene-copies per cell for diazotrophs in order to facilitate the necessary intercalibration of abundances between gene-counts and biomass. Zehr and Riemann (2023) point out that the compilation of Meiler et al. (2022) was too small to constrain the calibration tightly. It did, however, reflect the published information available at the time. Zehr and Riemann (2023) highlighted one avenue of progress by illustrating a recent data set (published subsequent to Meiler et al. 2022) that pleasingly reveals positive correlations between *nifH* gene and cell densities for specific diazotrophs in a recent field study (Gradoville et al. 2022). We look forward to more studies in which such relationships are examined across time and different ocean basins. The *nifH* gene abundance compilation (Tang

and Cassar 2019) available to Meiler et al. (2022) combined cyanobacterial diazotrophs into four groups, aggregating species with known variations in size, physiology, and biogeography and did not facilitate a finer taxonomically resolved comparison. To facilitate an appropriate data-model interface, it may be that simulations will need to resolve, or somehow parameterize, taxonomic resolution within the diazotrophs. Indeed, the biogeochemical model evaluated in Meiler et al. (2022) did resolve multiple size classes, but we note that almost all ocean biogeochemistry and climate models that explicitly represent diazotrophs represent them as a single population without any other trait or taxonomic resolution. Resolving more diversity is a computational challenge for climate models of the type used in global change simulations. Another potentially fruitful area for further work would be to better understand the value for quantitative evaluation of higher taxonomic resolution in both data and models. What is the trade-off between higher taxonomic resolution and data per type? Are there aggregating approaches which might exploit the taxonomic information without requiring explicit resolution of many classes in simulations?

Zehr and Riemann (2023) argue that the cell-to-biomass conversion ranges used by Meiler et al. (2022) inflate gene-to-biomass conversion errors when taxonomic groups of different cell sizes are combined. We agree that cell-count to biomass conversions are another important source of uncertainty, and this is also true when calibrating microscope- and flow-cytometer-based measures. Continued effort to catalog and calibrate cell-size in relevant groups is therefore valuable. We note also that most current biogeochemical models do not resolve the size spectrum (exceptions include Ward et al. 2013) in part because of computational costs making this even more challenging. Empirical calibration of relevant cell sizes, and appropriate size resolution in simulations represent two potentially valuable avenues for future effort.

Concluding thoughts

We appreciate the comments of Zehr and Riemann (2023). We agree that quantification of gene-copy number is valuable in marine microbial ecology, but we still contend that quantitative modeling of the global biogeography of diazotroph biomass from gene counts is currently subject to significant uncertainty. Zehr and Riemann's (2023) comments point us to several opportunities for future efforts, both empirical and theoretical, which can help reduce those uncertainties in intercalibration:

1. Additional gene copy per cell studies conducted both in the lab and in the field using consistent sampling protocols,
2. Comprehensive studies on biomass per cell of different taxonomic groups,
3. Continued and consistent measurements and reporting of *nifH* gene abundance resolving fine scales in taxonomy,
4. Efforts to better understand prevalence and mechanistic controls of polyploidy,
5. Improved taxonomic and allometric resolution of diazotrophs in biogeochemical models to appropriately interface with observations and optimize intercalibrations.

As outlined, there is a clear need to understand the distribution of diazotroph abundances and biomass to test biogeochemical models. As a group, marine diazotrophs are diverse taxonomically, functionally, and ecologically, with new groups still being discovered. Due to these biological differences, distributions of distinct groups of diazotrophs are differentially controlled by environmental factors, likely including those currently unknown. To construct a well-informed picture of global marine diazotroph biogeography going forward, studies need to continue to examine the ecology and distribution of different taxonomic diazotroph groups at fine scales, with qPCR being an important tool in this effort. We also suggest that in order to increase the confidence of comparisons across different *nifH* qPCR studies, efforts should continue on standardizing methods of *nifH* qPCR data collection and reporting. Further efforts to intercalibrate between currencies will be critical if we are to extract the maximum value for quantitative model testing, which we see as a major goal for global change modeling.

Finally, we reiterate that we agree *nifH* data have many valuable applications, including quantitative tests of biogeographical simulations and theory. Even with the large uncertainties, the Tang and Cassar (2019) data set provides clear constraints on the range of diazotrophy using presence–absence approach. This approach should be able to discriminate between the contradictory predictions of diazotroph biogeography found in CMIP5 climate models (Wrightson and Tagliabue 2020) as discussed above. The two classes of predicted biogeography cannot be simultaneously correct and the range constraint of presence–absence data should be sufficient to discriminate between them. With effort to minimize uncertainties through careful use of existing data, new calibration studies, and appropriate model formulations, evaluations of more subtle differences will also become possible.

Meiler et al. (2022) took a first step toward quantitative testing of modeled biogeography using a published *nifH* gene compilation. Zehr and Riemann's (2023) valuable comment inspires a roadmap toward reduced intercalibration uncertainties. We hope that this discussion will help focus research efforts toward elucidating the controls of marine diazotroph biogeography.

References

Aumont, O., and L. Bopp. 2006. Globalizing results from ocean in situ iron fertilization studies. Global Biogeochem. Cycl. **20**. doi:10.1029/2005GB002591

Biller, S. J., and others. 2018. Marine microbial metagenomes sampled across space and time. Sci. Data **5**: 180176. doi:10.1038/sdata.2018.176

Buitenhuis, E. T., and others. 2013. MAREDAT: Towards a world atlas of MARine Ecosystem DATa. Earth Syst. Sci. Data **5**: 227–239. doi:10.5194/essd-5-227-2013

Coles, V. J., and others. 2017. Ocean biogeochemistry modeled with emergent trait-based genomics. Science **358**: 1149–1154.

Doney, S. C., and others. 2009. Skill metrics for confronting global upper ocean ecosystem-biogeochemistry models against field and remote sensing data. J. Mar. Syst., **76**(1): 95–112. https://doi.org/10.1016/j.jmarsys.2008.05.015

Dunne, J. P., and others. 2013. GFDL's ESM2 global coupled climate–carbon earth system models. Part II: Carbon system formulation and baseline simulation characteristics. J. Climate **26**: 2247–2267. doi:10.1175/JCLI-D-12-00150.1

Dutkiewicz, S., B. A. Ward, F. Monteiro, and M. J. Follows. 2012. Interconnection of nitrogen fixers and iron in the Pacific Ocean: Theory and numerical simulations. Global Biogeochem. Cycles **26**: 1–16. doi:10.1029/2011GB004039

Follett, C. L., S. Dutkiewicz, D. M. Karl, K. Inomura, and M. J. Follows. 2018. Seasonal resource conditions favor a summertime increase in North Pacific diatom–diazotroph associations. ISME J. **12**: 1543–1557. doi:10.1038/s41396-017-0012-x

Friedrichs, M. A. M., and others. 2007. Assessment of skill and portability in regional marine biogeochemical models: Role of multiple planktonic groups. J. Geophys. Res. Oceans, **112**(C8): C08001.1–C08001.22.

Gradoville, M. R., M. Dugenne, A. M. Hynes, J. P. Zehr, and A. E. White. 2022. Empirical relationship between nifH gene abundance and diazotroph cell concentration in the North Pacific Subtropical Gyre. J. Phycol. **58**: 829–833. doi:10.1111/jpy.13289

Karl, D., R. Letelier, L. Tupas, J. Dore, J. Christian, and D. Hebel. 1997. The role of nitrogen fixation in biogeochemical cycling in the subtropical North Pacific Ocean. Nature **388**: 533–538. doi:10.1038/41474

Krupke, A., N. Musat, J. LaRoche, W. Mohr, B. M. Fuchs, R. I. Amann, M. M. M. Kuypers, and R. A. Foster. 2013. In situ identification and N_2 and C fixation rates of uncultivated cyanobacteria populations. Syst. Appl. Microbiol. **36**: 259–271. doi:10.1016/j.syapm.2013.02.002

Landolfi, A., W. Koeve, H. Dietze, P. Kähler, and A. Oschlies. 2015. A new perspective on environmental controls of marine nitrogen fixation. Geophys. Res. Lett. **42**: 4482–4489. doi:10.1002/2015GL063756

Luo, Y. W., and others. 2012. Database of diazotrophs in global ocean: Abundance, biomass and nitrogen fixation rates. Earth Syst. Sci. Data **4**: 47–73. doi:10.5194/essd-4-47-2012

Meiler, S., G. L. Britten, S. Dutkiewicz, M. R. Gradoville, P. H. Moisander, O. Jahn, and M. J. Follows. 2022. Constraining uncertainties of diazotroph biogeography from nifH gene abundance. Limnol. Oceanogr. **67**: 816–829. doi:10.1002/lno.12036

Monteiro, F. M., M. J. Follows, and S. Dutkiewicz. 2010. Distribution of diverse nitrogen fixers in the global ocean. Global Biogeochem. Cycl. **24**: 1–16. doi:10.1029/2009GB003731

Moore, C. M., and others. 2013. Processes and patterns of oceanic nutrient limitation. Nat. Geosci. **6**: 701–710. doi:10.1038/ngeo1765

Pesant, S., and others. 2015. Open science resources for the discovery and analysis of Tara Oceans data. Sci. Data **2**: 150023. doi:10.1038/sdata.2015.23

Ribalet, F., and others. 2019. SeaFlow data v1, high-resolution abundance, size and biomass of small phytoplankton in the North Pacific. Sci. Data **6**: 277. doi:10.1038/s41597-019-0292-2

Stukel, M. R., V. J. Coles, M. T. Brooks, and R. R. Hood. 2014. Top-down, bottom-up and physical controls on diatom-diazotroph assemblage growth in the Amazon River plume. Biogeosciences **11**: 3259–3278. doi:10.5194/bg-11-3259-2014

Tang, W., and N. Cassar. 2019. Data-driven modeling of the distribution of diazotrophs in the Global Ocean. Geophys. Res. Lett. **46**: 12258–12269. doi:10.1029/2019GL084376

Ward, B. A., S. Dutkiewicz, C. M. Moore, and M. J. Follows. 2013. Iron, phosphorus, and nitrogen supply ratios define the biogeography of nitrogen fixation. Limnol. Oceanogr. **58**: 2059–2075. doi:10.4319/lo.2013.58.6.2059

White, A. E., K. S. Watkins-Brandt, and M. J. Church. 2018. Temporal variability of *Trichodesmium* spp. and diatom-diazotroph assemblages in the North Pacific Subtropical Gyre. Front. Mar. Sci **5**: 1–12. doi:10.3389/fmars.2018.00027

Wrightson, L., and A. Tagliabue. 2020. Quantifying the impact of climate change on marine diazotrophy: Insights from earth system models. Front. Mar. Science **7**. doi:10.3389/fmars.2020.00635

Zehr, J., and L. Riemann. 2023. Quantification of gene copy numbers is valuable in marine microbial ecology; a comment to Meiler and others. (2022). Limnol. Oceanogr. **68**: 1406–1412. doi:10.1002/lno.12364

Acknowledgments

G.L.B. received support from the Simons Foundation Postdoctoral Fellowship, The Jessie B. Cox Charitable Trust, and the Woods Hole Oceanographic Institution. M.J.F. received funding from the Simons Foundation (SCOPE, Award 329108; CBIOMES, Award 549931). Open access funding provided by Eidgenossische Technische Hochschule Zurich.

Conflict of Interests

None declared.

Submitted 22 December 2022
Revised 30 March 2023
Accepted 18 April 2023

Associate editor: Ilana Berman-Frank