LIMNOLOGY and OCEANOGRAPHY

ASLO

Limnol. Oceanogr. 9999, 2023, 1–7

doi: 10.1002/lno.12364

Quantification of gene copy numbers is valuable in marine microbial ecology: A comment to Meiler et al. (2022)

Jonathan P. Zehr[1*] **Lasse Riemann**[2]

[1]Department of Ocean Sciences, University of California, Santa Cruz, California
[2]Department of Biology, Marine Biological Section, University of Copenhagen, Helsingør, Denmark

Marine nitrogen (N_2) fixation is important in the global biogeochemical cycling of N, and the interlinked cycling of carbon (C). Understanding and predicting global marine N_2 fixation requires information on diazotroph species specific rates of growth and N_2 fixation, their biogeography, and the physiology of nutrient limitation in diazotrophs and nondiazotrophs (for the basis of competition in models). Because many diazotrophs are not easily visualized or quantified and many are yet uncultivated, the application of polymerase chain reaction (PCR) methods to amplify the *nifH* gene, which encodes nitrogenase, the enzyme that catalyzes N_2 fixation, have led to multiple discoveries including new microorganisms (Zehr et al. 1998; Zehr and Capone 2020). Moreover, use of quantitative PCR (qPCR) has revealed unexpected biogeography of key marine diazotrophs regionally and globally (Church et al. 2005; Bentzon-Tilia et al. 2015; Langlois et al. 2015; Messer et al. 2015; Shiozaki et al. 2017; Harding et al. 2018; Mulholland et al. 2019). Biogeochemical modeling approaches have also been invaluable for providing hypothetical dynamic biogeography of the biomass associated with size classes of diazotrophs based on a number of assumptions (e.g., growth, nutrient uptake characteristics, and mortality; Dutkiewicz et al. 2014). It is, however, difficult to validate these models since there are few comprehensive datasets of N_2-fixing organism biogeography.

In a recent study, Meiler et al. (2022) attempted to use assembled qPCR *nifH* data to deduce the biogeography of key cyanobacterial diazotrophs using a derived currency of cell abundances and biomass, in order to assess the value of these data for validating numerical model biogeography. Using a few published studies (six publications that varied in species data and used different approaches) that had reported in situ data on *nifH* gene abundances, cell numbers or biomass (e. g. Hynes et al. 2012; Foster et al. 2013; Krupke et al. 2013; White et al. 2018) they derived conversion factors from *nifH* gene abundances to cell numbers, and from cells to biomass. The maximum and minimum range of derived conversion factors (i.e., the extremes, not average values) for both steps were then applied to an existing global marine database on *nifH* gene abundances (Luo et al. 2012; Tang and Cassar 2019). This approach generated ranges for depth integrated *nifH* gene abundance and *nifH* inferred cell counts and biomass across oceanic regions and latitudinal gradients for comparison with biogeochemical model output. At the biome level, the range in biomass predicted from *nifH* gene measurements was approximately four orders of magnitude greater than the dissimilarity in output from four model simulations from the same group over a period of 6 years. It was concluded that the error associated with converting *nifH* gene abundances to the model "currency" (form and units of variables needed by models) were too large for utility in model validation (ranging from four orders of magnitude for *Trichodesmium* to one order for *Crocosphaera*, UCYN-B). They concluded that qPCR data for the *nifH* gene, and likely other genes, were best used as an indicator of presence/absence rather than a measure of abundance for evaluating models. Although the analysis of Meiler et al. (2022) based on the commonly used approach of converting data to presence/absence is very useful and highlights the need for further research, their conclusion that "Despite its usefulness as an indicator of diazotrophic presence, *nifH* gene abundance may only weakly correlate with cell abundance and diazotrophic biomass" is worth revisiting.

We disagree with the conclusion that qPCR data are only weakly correlated with cell abundance because the analysis was based on very select few published datasets and (1) used assumptions about species and taxa biogeography based on size classes, (2) inflated the significance of polyploidy in undermining quantitation of diazotrophs, (3) used the resulting unnecessarily high ranges of combination of two conversion values as metrics, and (4) because the original *nifH* gene data reveal spatial patterns consistent with direct and indirect observations of diazotroph abundance and N_2 fixation rate contradicting the Meiler et al. (2022) assertion that *nifH* gene abundance is only of utility in presence/absence assessment.

*Correspondence: zehrj@ucsc.edu

Author Contribution Statement: Both authors contributed to the conceptualization and writing, and approved the final manuscript.

The problem with defined size classes for key diazotrophs

Meiler et al. (2022) argue that qPCR data are at the root of the currency disconnect between environmental measured abundances and model outputs, yet there are additional factors that affect abundance measurements based on qPCR (or any other molecular or imaging methods) and models. One important factor is that models use defined size classes, and many genera and species bridge different size classes and have different physiological characteristics that are not defined by size. The models define up to five size classes of diazotrophs between 3 and 15 μm in diameter (Dutkiewicz et al. 2021). The four cyanobacterial diazotrophs treated in the Meiler et al. (2022) study, *Trichodesmium*, the UCYN-A symbiosis, diatoms with *Richelia* symbionts, and *Crocosphaera* (some of which form large aggregates), vary in size across these size ranges (Pierella Karlusich et al. 2021). qPCR primers generally target a narrowly defined taxonomic group, and thus can distinguish between abundant and rare taxa, and define different sizes and biogeographic ranges. However, the data that were used combined taxa (qPCR data) within genus and species that varied not only in size, but also in frequency of occurrence in the oceans and biogeography.

The four groups addressed by Meiler et al. (2022) each have their own challenges for quantification. *Trichodesmium* spp. are a group of filamentous, cyanobacteria that do not form specialized N_2-fixing heterocyst cells (Zehr et al. 2023). *Trichodesmium* species vary greatly in size (6–16 μm in width and 4–23 μm long across very diverse species; Hynes et al. 2012), and other characteristics (Webb et al. 2009, Zehr et al. 2023; Hynes et al. 2012), but only a few species are typically abundant and found in wide areas of the ocean. Filamentous heterocyst-forming N_2-fixing cyanobacteria are associated with a number of genera of diatoms that also range in size and physical location of the symbiont, but also the number of filaments per diatom, the number of heterocysts per filament, and the degree of aggregation or chain formation (Caputo et al. 2019). UCYN-A (*Candidatus Atelocyanobacterium thalassa*) are a collection of mostly uncultivated strains of symbionts of the metabolically reduced cyanobacterium UCYN-A, endosymbiotic with the haptophyte *Braarudosphaera bigelowii* and relatives (Hagino et al. 2013; Farnelid et al. 2016). There are coastal and open ocean UCYN-A ecotypes that vary in size by at least a factor of 5 (Turk-Kubo et al. 2021). *Crocosphaera watsonii* is a marine planktonic species that is comprised of a number of very closely related strains (Webb et al. 2009; Zehr et al. 2023), include both free-living and symbiotic forms, but vary in size, physiology and aggregation. *C. watsonii* natural populations are comprised of large and small forms and aggregates (Zehr et al. 2023). Most common is the small form (Webb et al. 2009; Bench et al. 2016). Hence, as illustrated here (*see* also below), there is a fundamental currency problem when defining a size for diazotrophic genera or species, yet these different genera were all binned by Meiler et al. (2022). This problem inevitably contributes to variability in calculated *nifH* : cell and C : cell values across the different species and symbioses, but will also make any comparison between model output and *nifH* gene copy data highly questionable. There are qPCR primers for large and small *Crocosphaera*, different strains of *Trichodesmium*, *Richelia*/*Calothrix*, and UCYN-A, but Tang and Cassar (2019) combined the data for different strains/species within each group, which unfortunately was then used by Meiler et al. (2022) and contributed to the unnecessarily high ranges of conversion factors.

Number of nifH *genes per cell (polyploidy)*

The conclusions of Meiler et al. (2022) are based on analysis of a very small number of studies that reported in situ data for *nifH* gene abundance as well as some other measure of cell abundance or biomass (Krupke et al. 2013; Wilson et al. 2017; White et al. 2018) to calculate gene-to-cell conversion for the four major cyanobacterial groups based on the qPCR abundances of taxa-specific nitrogenase genes (*nifH*) and cell abundances. The studies differed in the target species, the specific measurements and methods used, and biogeographic location. Although these are among the few data available to attempt the desired conversion factors, the studies were not adequate for this purpose due to limitations that lead to erroneous conclusions regarding the ranges of values.

The studies cited used different approaches to determine *nifH*-to-cell conversions: White et al. (2018) used microscopy vs. qPCR; Wilson et al. (2017) used flow cytometry sorted cells coupled to qPCR; and Krupke et al. (2013) used catalyzed reporter in situ hybridization (CARD-FISH) vs. qPCR. These studies sometimes involved analyses from separate sample collections, as noted by Meiler et al. (2022), but this may be highly problematic for the conclusions drawn. For instance, *Trichodesmium* populations are comprised of free filaments and aggregates (puffs and tufts) that are buoyant and both float and sink and are notoriously heterogeneously distributed and difficult to sample (Rodier and Borgne 2008). Therefore, measurement comparisons need to be made on exactly the same water sample, not replicates from a CTD cast or replicate casts at the same station. Nonetheless, the cited study for *Trichodesmium* had an average *nifH* : cell that ranged by only a factor of 2 (130 $\pm$ 239 *nifH* : cell; White et al. 2018), compared to the total range Meiler et al. (2022) used for error propagation of 1.4–1405 *nifH* per cell. Furthermore, some of the studies also compared different sample types, such as different size-fractionations. For example, Krupke et al. (2013) used different samples (pooled FISH filter size fractions to compare to 0.2-μm filtered DNA samples) to measure abundances of UCYN-A by CARD-FISH and *nifH* gene abundance by qPCR, respectively, thus involving another source of error for comparison purposes. A recent study shows that *nifH* gene

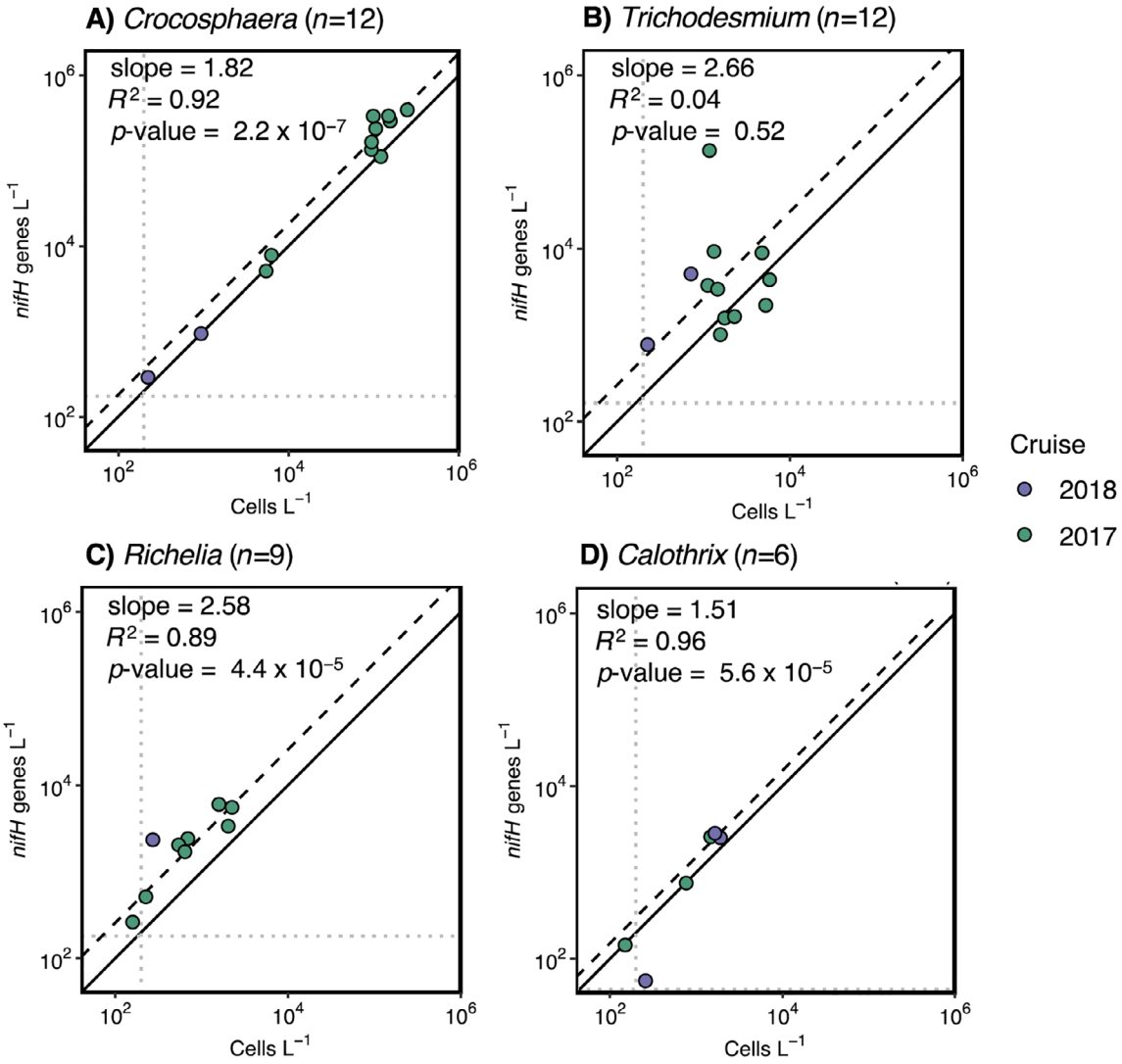

Fig. 1. Correlation of *nifH* gene abundances vs. diazotroph cell concentrations from two cruises in 2017 and 2018. Solid lines show a 1 : 1 relationship; dashed lines show the slope from a simple linear regression model with a fixed zero intercept. Regression statistics are provided for each subplot. Gray dotted lines reflect detection limits, corresponding to 50 cells L^{-1} for 20 mL binned samples (~ 200 cells L^{-1}) for IFCb measurements and 44–181 *nifH* genes L^{-1} for ddPCR measurements (Gradoville et al. 2021). Note that data are plotted on a logarithmic scale but regressions were performed using untransformed data. From Gradoville et al. (2022), with permission.

abundance can be well-correlated, certainly not varying by orders of magnitude (Fig. 1, panels A-D, Gradoville et al. 2022).

One of the fundamental limitations of the approach taken by Meiler et al. (2022) is due to the aggregation of species/strains that are known to vary greatly in size (or sizes of symbiotic hosts), physiology and biogeography, which unfortunately were combined in the Tang and Cassar (2019) database used for the study. For example, data were combined from individual measurements for UCYN-A1 and UCYN-A2, that differ in cell size and biomass, biogeography, and physiology (Farnelid et al. 2016; Turk-Kubo et al. 2021) to represent UCYN-A in the models that address the global open ocean. Similarly, the diatom symbiont species were combined, despite the fact that there are qPCR probes that target the individual symbioses that differ in size, physiology, and biogeography. In the case of *Richelia* there are a number of sources of variability of *nifH* : cell and cell : C, as *Richelia* and related strains form associations with several genera of diatoms of varying sizes, with differing numbers of symbiotic filaments and numbers of cells per filament (Pierella Karlusich et al. 2021; Flores et al. 2022).

A variety of issues pertain to quantifying cell numbers from any gene (DNA) measurement. Since many of the diazotroph species targeted by qPCR are uncultivated (e.g., UCYN-A1), there are not precise estimates of extraction efficiencies as

noted by Meiler et al. (2022), but they are likely relatively constant within species. Furthermore, extraction efficiency is also an issue with any DNA based approach, such as metagenomics (Pierella Karlusich et al. 2021), so the implications are not unique to qPCR. Polyploidy, or multiple genome copies per cell, has long been known in cyanobacteria (and eukaryotes), but only recently recognized with respect to N_2-fixing cyanobacteria in oceans (Sargent et al. 2016; Pierella Karlusich et al. 2021). Although little is known, as there have been few direct studies, the potential for polyploidy to affect cell abundance estimates may be significant. However, even for *Trichodesmium* field data suggest that the average copy number could be easily modeled based on the linear relationship of cell number vs. *nifH* gene copy number (Fig. 2). Thus, a more straightforward approach, until there are further defining studies, would be to make simple assumptions about the relationship between cell and *nifH* gene copies for common representative species, as exemplified for *Trichodesmium* (Fig. 2). Nonetheless, we do advocate that extreme caution should be taken when converting gene copies to cell number, and in many cases this conversion is not necessary or desirable (*see* below).

The example data provided in Figs. 1, 2 show that qPCR data can be related to cell numbers and certainly are not consistent with the wide range of conversion factors for *nifH* gene copy : cell used by Meiler et al. (2022). Rather, we find that the choice by Meiler et al. (2022) of using range as the metric amplifies the currency problems and that the problems are not specifically related to qPCR per se (e.g., the biomass conversion problem would also apply to cell counts).

Cell-to-biomass conversions inflate the qPCR: Biomass error

The cell-to-biomass conversion calculations are perhaps the most problematic analysis in Meiler et al. (2022). In order to convert cell numbers to biomass (in C units), Meiler et al. (2022) used a few published datasets (Foster et al. 2011; Hynes et al. 2012; Wilson et al. 2017; Harding et al. 2018) and biomass : carbon conversion parameters (Strathman 1967; Verity et al. 1992). The estimated biomass is very sensitive to the measurements of size, assumed dimensions, and assumed elemental stoichiometry.

Only a few of the studies directly measured cell C on cultures (Strathman 1967; Verity et al. 1992). Much of the cell C data came from studies that were not designed to directly determine universal cell : C conversion factors, but were estimating cell C in order to determine isotope enrichment in labeled C and N uptake studies by nanoscale secondary ion mass spectrometry (nanoSIMS; Foster et al. 2011; Krupke et al. 2013; Harding et al. 2018). In nanoSIMS, cells are only partially analyzed as an ion beam ionizes successive layers of the cell, and the total biomass measured is dependent on the orientation of the cell and the fraction of the cell "burned" through. Total cell C is estimated by calculating volumes based on assumed 3D shapes from two-dimensional slice data. Although this approach is necessary for nanoSIMS isotope tracer experimental analysis, accurate estimates of cellular C are better derived from cultivated organisms and/or other techniques. Furthermore, in the case of UCYN-A, the study chosen for analysis is from the extremes of the biogeographic ranges for UCYN-A (the Arctic; Harding et al. 2018), which is not the best

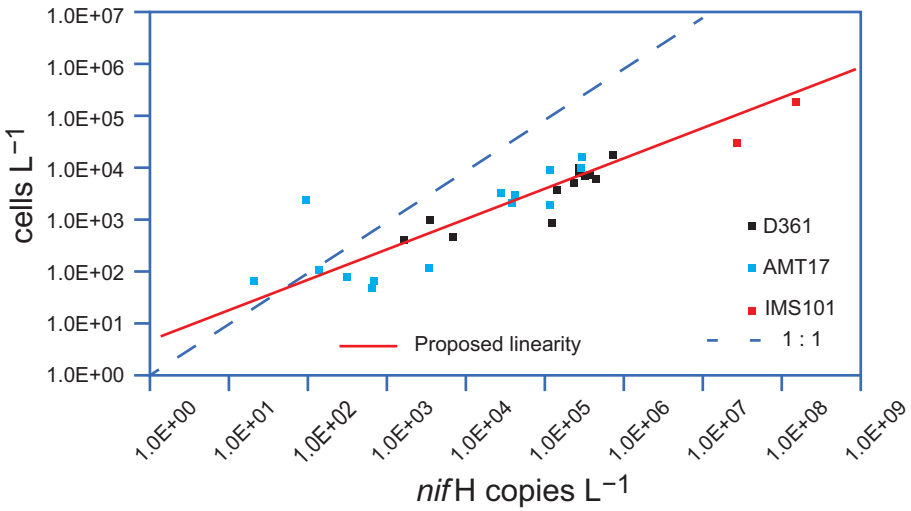

Fig. 2. Relationship between *Trichodesmium* abundance and *nifH* gene copy number obtained from surface samples on two transects (ATM17, blue and D361, black) in the tropical and subtropical Atlantic Ocean. *Trichodesmium erythraeum* IMS101 culture samples are shown in red. Proposed linear relationship drawn in to indicate relationship between gene copy and cell numbers. Redrawn from Sargent et al. (2016), with permission.

representative for the species, particularly for the open ocean that is the focus of the models.

The range of cell : C conversions is amplified by combining strains/species of known differences in size and biogeography. For example, Meiler et al. (2022) combined UCYN-A strains, even though the strain that inhabits the open ocean, UCYN-A1, is much smaller than the others and can be constrained in size. The size of the UCYN-A1 host, which is the relevant measure for modeling biomass, is only 2–3 μm in diameter, whereas the coastal UCYN-A2 strains can be over five times this size (Cabello et al. 2016).

The limitations of cell : C conversions apply when calculating biomass numbers from any cell data to biomass C or N, which are the typical currency of models, independent of the method used to derive the conversions (including microscopy or flow cytometry). Thus, such conversions do not apply only to qPCR data, but represent a general challenge in a wide variety of environmental studies that seek to extrapolate specific measurements to their biogeochemical impact at larger scales, including modeling approaches (Coles et al. 2017).

Diazotroph qPCR data are valuable

Meiler et al. (2022) describe the known limitations of qPCR data, including the issue of polyploidy and sources of variance associated with any DNA (or RNA) based methods. Combined with their handling of conversion factors and choice of range as a metric, this may leave the casual reader with the impression that qPCR is a flawed technique with no value. However, qPCR data can allow for quantitative comparisons in abundances even if there are limitations in converting to cell number and biomass. We stress that qPCR data have demonstrated important characteristics of the biogeography of diazotrophs, including seasonality (Church et al. 2009; Bentzon-Tilia et al. 2015; Cabello et al. 2016; Tang and Cassar 2019), biogeography (Shiozaki et al. 2017; Hallstrøm et al. 2022) and hotspots or blooms of uncultivated or difficult to detect organisms (Goebel et al. 2008; Moisander et al. 2010). This has led to invaluable insights into the environmental drivers of diazotrophic taxa now known to be of key importance to marine N cycling. Also, high abundances measured by qPCR made it possible to define samples for cell sorting and genome sequencing of UCYN-A (Bombar et al. 2014). qPCR is the only current method allowing quantification of uncultivated diazotrophs at sea, which is critical for obtaining more data for defining biogeography and validating models given the large spatial and temporal scales and patchiness of natural populations.

Finally, if we examine the *nifH* dataset used by Meiler et al. (2022), we find that in fact, not only do the spatial patterns of gene abundance closely match independent estimates of diazotroph biomass, but they echo the distribution of N_2 fixation rates from compilations of observations, inverse models based on nutrient distributions, and ecosystem models (Luo et al. 2012; Aumont et al. 2015; Landolfi et al. 2015; Letscher et al. 2015; Wang et al. 2019). Indeed, the analysis of Luo et al. (2012) while lacking a number of newer samples in the more recent Tang and Cassar (2019) compilation, demonstrates the close relationship between depth integrated diazotroph carbon biomass (their fig. 8a) and depth integrated *nifH* counts converted to carbon biomass (their fig. 9a).

Summary

The main conclusion of Meiler et al. (2022) is that caution should be taken when extrapolating from gene counts (qPCR, but also applies to metagenomic data) to cells or biomasses of organisms. This is an important point, but we argue that the ranges provided for gene : cell or cell : biomass relationships are poorly supported by current data. Moreover, we find that using such range extrema to evaluate empirical observations is inappropriate and unintentionally support the conclusion that gene counts are a poor basis for understanding controls of marine diazotrophy.

Taken together, we find that the conversion factors calculated by Meiler et al. (2022) are based on too few data and that these in several cases are obtained with variable methods and sampling strategies unsuited for calculating gene: cell or cell: biomass relationships. Hence, we find that (1) the confidence in the ranges of conversion factors calculated by Meiler et al. (2022), which forms the backbone of their study, is limited, which makes their conclusion on the restricted utility of qPCR data questionable; (2) their study emphasizes the importance of ranges, although methods typically use reasonable estimates or averages rather than ranges that represent the extremes; and (3) the basic premise for their study, the conversion of genes to cell abundance and biomass, is constrained to the purpose of integrating gene counts into biogeochemical models, whereas most studies using qPCR have a different goal.

In summary, there are several implications of the Meiler et al. (2022) study. First, the study combines errors associated with DNA measures with those of cell biomass. This second source of error, ranging over two orders of magnitude in Meiler et al. (2022), does not only apply to qPCR studies but to many modeling applications using many sources of data for cell abundance and thus does not apply directly to qPCR data but to virtually all oceanographic data. The discussion of Meiler et al. (2022) leads to the implication that there is no quantitative value in the spatial patterns observed, yet the extant patterns do agree well with what is known about diazotroph N_2 fixation biogeography. More generally, if gene abundance does not reflect cellular genome content, then many microbial oceanographic studies, not involving modeling, would need to be re-evaluated. In conclusion, the Meiler et al. (2022) study provided important and thought-provoking insights, but it should be noted that qPCR data, of other genes as well as *nifH*, are quantitative and valuable. We encourage

continued use of qPCR enumeration of key genes to gain insights into the biogeography, controls, and ecology of microbes in the oceans.

References

Aumont, O., C. Ethé, A. Tagliabue, L. Bopp, and M. Gehlen. 2015. PISCES-v2: An ocean biogeochemical model for carbon and ecosystem studies. Geosci. Model Dev. **8**: 2465–2513. doi:10.5194/gmd-8-2465-2015

Bench, S. R., I. Frank, J. Robidart, and J. P. Zehr. 2016. Two subpopulations of *Crocosphaera watsonii* have distinct distributions in the North and South Pacific. Environ. Microbiol. **18**: 514–524.

Bentzon-Tilia, M., and others. 2015. Significant N_2 fixation by heterotrophs, photoheterotrophs and heterocystous cyanobacteria in two temperate estuaries. ISME J. **9**: 273–285.

Bombar, D., P. Heller, P. Sanchez-Baracaldo, B. J. Carter, and J. P. Zehr. 2014. Comparative genomics reveals surprising divergence of two closely related strains of uncultivated UCYN-A cyanobacteria. ISME J. **8**: 2530–2542.

Cabello, A. M., and others. 2016. Global distribution and vertical patterns of a prymnesiophyte-cyanobacteria obligate symbiosis. ISME J. **10**: 693–706. doi:10.1038/ismej.2015.147

Caputo, A., J. A. Nylander, and R. A. Foster. 2019. The genetic diversity and evolution of diatom-diazotroph associations highlights traits favoring symbiont integration. FEMS Microbiol. Lett. **366**: fny297. doi:10.1093/femsle/fnz120

Church, M., B. Jenkins, D. Karl, and J. Zehr. 2005. Vertical distributions of nitrogen-fixing phylotypes at Stn ALOHA in the oligotrophic North Pacific Ocean. Aquat. Microb. Ecol. **38**: 3–14.

Church, M. J., C. Mahaffey, R. M. Letelier, R. Lukas, J. P. Zehr, and D. M. Karl. 2009. Physical forcing of nitrogen fixation and diazotroph community structure in the North Pacific subtropical gyre. Global Biogeochem. Cycles **23**: GB2020. doi:10.1029/2008GB003418

Coles, V., and others. 2017. Ocean biogeochemistry modeled with emergent trait-based genomics. Science **358**: 1149–1154. doi:10.1126/science.aan5712

Dutkiewicz, S., B. A. Ward, J. R. Scott, and M. J. Follows. 2014. Understanding predicted shifts in diazotroph biogeography using resource competition theory. Biogeosciences **11**: 5445–5461.

Dutkiewicz, S., P. W. Boyd, and U. Riebesell. 2021. Exploring biogeochemical and ecological redundancy in phytoplankton communities in the global ocean. Glob. Chang. Biol. **27**: 1196–1213. doi:10.1111/gcb.15493

Farnelid, H., K. Turk-Kubo, M. C. Muñoz-Marín, and J. P. Zehr. 2016. New insights into the ecology of the globally significant uncultured nitrogen-fixing symbiont UCYN-A. Aquat. Microb. Ecol. **77**: 125–138.

Flores, E., D. K. Romanovicz, M. Nieves-Morión, R. A. Foster, and T. A. Villareal. 2022. Adaptation to an intracellular lifestyle by a nitrogen-fixing, heterocyst-forming cyanobacterial endosymbiont of a diatom. Front. Microbiol. **13**. doi:10.3389/fmicb.2022.799362

Foster, R. A., M. M. M. Kuypers, T. Vagner, R. W. Paerl, N. Musat, and J. P. Zehr. 2011. Nitrogen fixation and transfer in open ocean diatom-cyanobacterial symbioses. ISME J. **5**: 1484–1493. doi:10.1038/ismej.2011.26

Foster, R. A., S. Sztejrenszus, and M. M. Kuypers. 2013. Measuring carbon and N_2 fixation in field populations of colonial and free-living unicellular cyanobacteria using nanometer-scale secondary ion mass spectrometry. J. Phycol. **49**: 502–516. doi:10.1111/jpy.12057

Goebel, N. L., C. A. Edwards, B. J. Carter, K. M. Achilles, M. J. Church, and J. P. Zehr. 2008. Growth and carbon content of three different sized diazotrophic cyanobacteria observed in the subtropical North Pacific. J. Phycol. **44**: 1212–1220.

Gradoville, M. R., A. M. Cabello, S. T. Wilson, K. A. Turk-Kubo, D. M. Karl, and J. P. Zehr. 2021. Light and depth dependency of nitrogen fixation by the non-photosynthetic, symbiotic cyanobacterium UCYN-A. Environ. Microbiol. **23**: 4518–4531.

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.

Hagino, K., R. Onuma, M. Kawachi, and T. Horiguchi. 2013. Discovery of an endosymbiotic nitrogen-fixing cyanobacterium UCYN-A in *Braarudosphaera bigelowii* (Prymnesiophyceae). PloS One **8**: e81749. doi:10.1371/journal.pone.0081749

Hallstrøm, S., M. Benavides, E. R. Salamon, J. Arístegui, and L. Riemann. 2022. Activity and distribution of diazotrophic communities across the Cape Verde Frontal Zone in the Northeast Atlantic Ocean. Biogeochemistry **160**: 49–67.

Harding, K., K. A. Turk-Kubo, R. E. Sipler, M. M. Mills, D. A. Bronk, and J. P. Zehr. 2018. Symbiotic unicellular cyanobacteria fix nitrogen in the Arctic Ocean. Proc. Natl. Acad. Sci. U.S.A. **115**: 13371–13375. doi:10.1073/pnas.1813658115

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

Krupke, A., and others. 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.

Langlois, R., T. Großkopf, M. Mills, S. Takeda, and J. LaRoche. 2015. Widespread distribution and expression of gamma a

(UMB), an uncultured, diazotrophic, γ-proteobacterial *nifH* phylotype. PloS One **10**: e0128912. doi:10.1371/journal.pone.0128912

Letscher, R. T., J. K. Moore, Y. C. Teng, and F. Primeau. 2015. Variable C:N:P stoichiometry of dissolved organic matter cycling in the Community Earth System Model. Biogeosciences **12**: 209–221. doi:10.5194/bg-12-209-2015

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., and others. 2022. Constraining uncertainties of diazotroph biogeography from *nifH* gene abundance. Limnol. Oceanogr. **67**: 816–829. doi:10.1002/lno.12036

Messer, L. F., and others. 2015. High levels of heterogeneity in diazotroph diversity and activity within a putative hotspot for marine nitrogen fixation. ISME J. **10**: 1499–1513. doi:10.1038/ismej.2015.205

Moisander, P. H., and others. 2010. Unicellular cyanobacterial distributions broaden the oceanic N_2 fixation domain. Science **327**: 1512–1514. doi:10.1126/science.1185468

Mulholland, M. R., and others. 2019. High rates of N_2 fixation in temperate, western North Atlantic coastal waters expands the realm of marine diazotrophy. Global Biogeochem. Cycl. **33**: 826–840. doi:10.1029/2018GB006130

Pierella Karlusich, J. J., and others. 2021. Global distribution patterns of marine nitrogen-fixers by imaging and molecular methods. Nat. Comm. **12**: 4160. doi:10.1038/s41467-021-24299-y

Rodier, M., and R. Borgne. 2008. Population dynamics and environmental conditions affecting *Trichodesmium* spp. (filamentous cyanobacteria) blooms in the south-west lagoon of New Caledonia. J. Exp. Mar. Biol. Ecol. **358**: 20–32.

Sargent, E. C., and others. 2016. Evidence for polyploidy in the globally important diazotroph *Trichodesmium*. FEMS Microbiol. Lett. **363**: fnw244. doi:10.1093/femsle/fnw244

Shiozaki, T., and others. 2017. Basin scale variability of active diazotrophs and nitrogen fixation in the North Pacific, from the tropics to the subarctic Bering Sea. Global Biogeochem. Cycl. **31**: 996–1009. doi:10.1002/2017gb005681

Strathman, R. 1967. Estimating the organic carbon content of phytoplankton from cell volume or plasma volume. Limnol. Oceanogr. **12**: 411–418. doi:10.4319/lo.1967.12.3.0411

Tang, W., and N. Cassar. 2019. Data-driven modeling of the distribution of diazotrophs in the global ocean. Geophys. Res. Lett. **46**: 12258–12269.

Turk-Kubo, K. A., and others. 2021. UCYN-A/haptophyte symbioses dominate N_2 fixation in the Southern California Current System. ISME Comm. **1**: 42. doi:10.1038/s43705-021-00039-7

Verity, P. G., C. Y. Robertson, C. R. Tronzo, M. G. Andrews, J. R. Nelson, and M. E. Sieracki. 1992. Relationships between cell volume and the carbon and nitrogen content of marine photosynthetic nanoplankton. Limnol. Oceanogr. **37**: 1434–1446. doi:10.4319/lo.1992.37.7.1434

Wang, W. L., J. K. Moore, A. C. Martiny, and F. W. Primeau. 2019. Convergent estimates of marine nitrogen fixation. Nature **566**: 205–211. doi:10.1038/s41586-019-0911-2

Webb, E. A., I. M. Ehrenreich, S. L. Brown, F. W. Valois, and J. B. Waterbury. 2009. Phenotypic and genotypic characterization of multiple strains of the diazotrophic cyanobacterium, *Crocosphaera watsonii*, isolated from the open ocean. Environ. Microbiol. **11**: 338–348. doi:10.1111/j.1462-2920.2008.01771.x

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**. doi:10.3389/fmars.2018.00027

Wilson, S. T., and others. 2017. Coordinated regulation of growth, activity and transcription in natural populations of the unicellular nitrogen-fixing cyanobacterium *Crocosphaera*. Nat. Microbiol. **2**: 17118.

Zehr, J., M. Mellon, and S. Zani. 1998. New nitrogen-fixing microorganisms detected in oligotrophic oceans by amplification of nitrogenase (*nifH*) genes. Appl. Environ. Microbiol. **64**: 3444–3450. doi:10.1128/AEM.64.9.3444-3450.1998

Zehr, J. P., and D. G. Capone. 2020. Changing perspectives in marine nitrogen fixation. Science **368**: eaay9514.

Zehr, J. P., R. A. Foster, J. Waterbury, and E. A. Webb. 2023. *Crocosphaera*. *In* M. E. Trujillo, S. Dedysh, P. DeVos, B. Hedlund, P. Kämpfer, F. A. Rainey, and W. B. Whitman [eds.], Bergey's manual of systematics of archaea and bacteria. doi:10.1002/9781118960608.gbm01517

Acknowledgments

The authors thank Kendra Turk-Kubo, Mary R. Gradoville, Victoria J. Coles, Rachel Foster, Annette Hynes, Eric Webb, Tracy Villareal, Sam Wilson, and Angelicque White for suggestions and insights. J.Z. was supported by the Simons Foundation (SCOPE #724220 and Life Sciences #824082). L.R. was supported by grant 0217-00089B from Independent Research Fund Denmark.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Submitted 19 August 2022
Revised 18 October 2022
Accepted 04 November 2022

Associate editor: K. David Hambright