# A Short Comparison of Two Marine Planktonic Diazotrophic Symbioses Highlights an Un-quantified Disparity

Andrea Caputo\*, Marcus Stenegren, Massimo C. Pernice and Rachel A. Foster

*Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm, Sweden*

Some $N_2$-fixing cyanobacteria form symbiosis with diverse protists. In the plankton two groups of diazotrophic symbioses are described: (1) a collective group of diatoms which associate with heterocystous cyanobacteria (Diatom Diazotroph Associations, DDA), and (2) the microalgal prymnesiophyte *Braarudosphaera bigelowii* and its relatives which associate with the unicellular cyanobacterium *Candidatus* Atelocyanobacterium thalassa (hereafter as UCYN-A). Both symbiotic systems co-occur, and in both partnerships the symbionts function as a nitrogen (N) source. In this perspective, we provide a brief comparison between the DDAs and the prymnesiophyte-UCYN-A symbioses highlighting similarities and differences in both systems, and present a bias in the attention and current methodology that has led to an under-detection and under-estimation of the DDAs.

Keywords: cyanobacteria, diazotrophs, diatoms, symbiosis, DDA, UCYN-A, *Richelia*, *Calothrix*

## INTRODUCTION

Some of the most inconspicuous components of the plankton are partnerships, or symbioses, between diverse unicellular eukaryotic microalgae and prokaryotic cyanobacteria. In some planktonic symbiosis, the symbionts are $N_2$ fixers, or diazotrophs, that reduce di-nitrogen ($N_2$) and provide ammonia to their respective hosts. There are two marine planktonic $N_2$ fixing symbiotic systems described, which can be divided by host type. The more recently discovered is a partnership between a prymnesiophyte, the microalga *Braarudosphaera bigelowii* and its relatives and several lineages of the unicellular cyanobacterium, *Candidatus* Atelocyanobacterium thalassa or more commonly called UCYN-A (Zehr et al., 2008; Tripp et al., 2010; Thompson et al., 2012). The second symbiotic system is a group referred to as diatom diazotroph associations (DDAs) (Foster and O'Mullan, 2008).

DDAs are symbioses between various diatom genera (*Rhizosolenia, Hemiaulus, Chaetoceros, Guinardia, Bacteriastrum*) and the heterocystous cyanobacteria, *Richelia intracellularis* or *Calothrix rhizosoleniae*. Each symbiotic filament is composed of several vegetative cells and a terminal heterocyst for $N_2$ fixation. There is an additional DDA described between the pennate diatom, *Climacodium frauenfeldianum*, and a unicellular cyanobacteria (Carpenter and Janson, 2000). For the purposes of this review, however, we limit the discussion of the DDAs to the partnerships involving heterocystous cyanobacterial symbionts.

Planktonic $N_2$ fixing symbioses are widespread, often co-occur and both are considered significant contributors to both the N and C cycles. The DDAs were described over a century ago, yet remain understudied, with only two dedicated reviews in the last two decades (Villareal, 1992; Foster and O'Mullan, 2008), while the prymnesiophyte-UCYN-A symbioses were recently

discovered (within a decade) and have been highlighted in several reviews (e.g., Bothe et al., 2010; Delong, 2010; Thompson and Zehr, 2013; Zehr, 2015; Farnelid et al., 2016; Zehr et al., 2016). Moreover, rarely are the two systems directly compared despite their similar nature as planktonic symbioses and function in the marine N cycle. The purpose of this short perspective is to provide a brief comparison of these two biogeochemically relevant symbioses and highlight a current disparity in attention and detection for the UCYN-A based symbiosis.

## EARLY OBSERVATIONS AND FIRST GENETIC IDENTITY STUDIES

The earliest reports of DDAs came from simple microscopic observations in the late nineteenth and early twentieth centuries by Lemmermann (1905) and Ostenfeld and Schmidt (1901) when they observed and described the cyanobacterial epiphytes on *Rhizosolenia, Hemiaulus,* and *Chaetoceros* spp. diatoms. For over a century the DDAs were presumed as N₂ fixing symbiosis, since it was only recently shown that N was fixed and transferred to the hosts (Foster et al., 2011). While, the UCYN-A was initially believed as free-living and was only recently identified as a symbiont (Thompson et al., 2012). Unlike the DDAs which can be distinguished by standard epi-fluorescent microscopy, the UCYN-A symbiosis require sophisticated (e.g., competitor and dual labeled) and highly specific fluorescent *in situ* hybridization (FISH) based assays (Thompson et al., 2012; Krupke et al., 2013; Cabello et al., 2016; Cornejo-Castillo et al., 2016). However the earliest detection of UCYN-A (and DDA symbionts) were by environmental *nifH* sequence (encodes for nitrogenase enzyme complex for N₂ fixation) libraries.

In the early diversity studies using the *nifH* gene, the UCYN-A lineage and two of the three heterocystous lineages for DDAs were identified (Zehr et al., 1998, 2001; Church et al., 2005). These first sequences were referred to as: group A (hereafter UCYN-A), het-1, and het-2, however no sequence was directly linked to a particular cell or symbiosis. The het-1 and het-2 *nifH* sequences were similar to *nifH* sequences of other heterocystous cyanobacteria and therefore assumed to be derived from the symbionts of diatoms since no other heterocystous cyanobacteria were known to exist in the open ocean freely.

The identity of UYCN-A as a symbiont remained unknown for another decade, while the first two het lineages (het-1, het-2) and a third (het-3) were soon after identified as symbionts of *Rhizosolenia* spp*., Hemiaulus* spp., *Chaetoceros* spp. diatoms, respectively (Foster and Zehr, 2006). The het-3 symbiont of *Chaetoceros compressus* was isolated into culture, genome sequenced (CalSC01) and has been maintained asymbiotic since 2004 (Foster et al., 2010; Hilton et al., 2013). Currently there are six *nifH* clades designated as UCYN-A (UCYNA-1 through UCYN-A6), two clades are derived from symbiotic cells (i.e., UCYN-A1, UCYN-A2) (Thompson et al., 2014; Farnelid et al., 2016; Turk-kubo et al., 2017) and no UCYN-A cells have been isolated into culture.

## SPECIFICITY, TRANSMISSION, AND MAINTENANCE

The DDAs are described as highly host specific, such that one symbiont type (e.g., het-1) associates with one host genus (*Rhizosolenia*) (Janson et al., 1999; Foster and Zehr, 2006). In fact, the specific nature in the partnerships allows the various DDAs to be quantified by quantitative PCR (qPCR) assays specific for the *nifH* sequence of the symbiont (see below). Similar high host fidelity is described in the two UCYN-A symbioses (Thompson et al., 2014; Cabello et al., 2016). The driver of specificity is unknown and unstudied in both symbiotic systems. Since both systems have evaded cultivation either entirely (e.g., UCYN-A) or for long-term (e.g., DDAs) our understanding of the life cycles and host-symbiont communication is extremely limited.

The infection and transmission (e.g., horizontal or vertical) has not been observed in either system. Earlier laboratory studies on the *Rhizosolenia-Richelia* symbioses found asynchronicity between partner growth rates and division cycles, which led to asymbiotic hosts (Villareal, 1990). Observations of asymbiotic diatom hosts and/or free-living *R. intracellularis* and *C. rhizosoleniae* are less common in the literature (Sournia, 1970; Marumo and Asaoka, 1974; Villareal, 1990; Gómez et al., 2005; White et al., 2007). Equally unknown is if another life stage persists or is required for infection in either symbioses as has been reported in terrestrial based symbioses with cyanobacteria (e.g., hormogonia) (Rasmussen, 2002). Given the rare instances of UCYN-A in asymbiotic form by CARD-FISH studies (Thompson et al., 2012; Krupke et al., 2013; Cabello et al., 2016), it is unlikely that another life stage persists for UCYN-A.

## SYMBIONT LOCATION

A major finding when the draft genomes for the *Richelia* (RintHH01) and *Calothrix* (CalSC01) symbionts were reported was that the symbiont genome size and dependency was linked to location in the host diatoms (Hilton et al., 2013). A closer look at the two symbioses reveals that the housing of the symbionts are quite similar. For example, in prymnesiophyte hosts, UCYN-A have been described as internal symbionts by TEM observations of the host cytoplasm, loosely attached or in a free-floating compartment (symbiosome) (Hagino et al., 2013; Cornejo-Castillo et al., 2016). The symbionts of the DDAs vary from externally attached to a quasi-internal location residing between the host frustule and plasma membrane (Villareal, 1992). Recently using confocal microscopy we observed evidence for an internal location of *R. intracellularis* in the cytoplasm of *Hemiaulus hauckii* diatoms (Caputo et al. unpubl.). Thus the location for the symbionts in the two systems varies: external, quasi-, loosely attached, and internal.

## SYMBIONT GENOME COMPARISON

Great insight into UCYN-A potential as a symbiont was revealed by sequencing and closing its genome (Zehr et al., 2008; Tripp et al., 2010). Likewise the sequencing of the draft genomes for

the DDA symbionts (RintHH01, RintHM01, CalSC01) revealed a remarkable reduction in genome size and content in the internal symbiont compared to the external symbiont (Hilton et al., 2013). There is an additional draft *Richelia* genome (RintRC01) now available (Hilton, 2014). Two UCYN-A genomes are available, one is closed, and a second is open (Zehr et al., 2008; Tripp et al., 2010; Bombar et al., 2014). We excluded the DDA symbiont RintHM01 as it was reduced and lacked genes expected for a full genome (Hilton, 2014). Here we compare all symbiont genomes by quantifying the number of genes encoding for a particular subsystem normalized by genome size (Supplementary Method).

The genome size for the symbionts are similar in that all but the external DDA (CalSC01) have been reduced compared to free-living cyanobacteria. The deleted genome content in the DDA and UCYN-A symbionts however differs and was previously highlighted (**Table 1**; Zehr et al., 2008; Tripp et al., 2010; Hilton et al., 2013). An interesting similarity, which was not reported, was a reduction in the number of genes for $CO_2$ uptake in the RintHH01 (internal symbiont of diatom host *H. hauckii*). In fact the decrease in genes for $CO_2$ uptake seems to be balanced by an increase in number of genes for $N_2$ fixation. Thus similar to UCYN-A1 and A2, the RintHH01 appears to be reducing its capacity for $CO_2$ uptake, while investing in $N_2$ fixation.

The number of genes encoding for K homeostasis and P uptake are similar in the DDA and the UCYN-A symbionts and perhaps related to occurrences in similarly saline and low P environments. The genomes for the UCYN-A symbionts contain more genes for $N_2$ fixation, vitamin synthesis, and cell division, compared to the three DDA symbionts. An increase in genes for $N_2$ fixation is consistent with recent studies showing a streamlined genome expression toward $N_2$ fixation (Cornejo-Castillo et al., 2016). The role of vitamins in the two symbiotic systems is un-studied and of interest given that at least half of the algal kingdom requires an exogenous supply of vitamin $B_{12}$ and a large percentage of phytoplankton, including prymnesiophytes (80%) are vitamin $B_1$ auxotrophs (Croft et al., 2005, 2006).

Both the external (CalSC01) and quasi-internal (RintRC01) DDA symbiont draft genomes contain more genes for transport than the internal residing symbiont (RintHH01) (**Figure 1**, **Table 1**). The number of genes for transport in UCYN-A symbionts is similar to CalSC01 and is coherent with a higher investment of the genome for transport given an external location (**Table 1**). The internal DDA symbiont (RintHH01) and UCYN-A symbionts contain similar numbers of genes for stress, including genes encoding for oxidative stress.

In summary, there were more parallels in the symbiont genomes than previously highlighted. As reported earlier, the genomes of UCYN-A1 and UCYN-A2 are similar in size, content, and streamlining (Bombar et al., 2014); while the DDA symbiont draft genomes differ largely (**Figure 1**, **Table 1**), and content appears influenced by the symbiont location (Hilton et al., 2013).

# METHODOLOGICAL BIASES: UN-QUANTIFIED AND UNDER-SAMPLED DISPARITIES

Prior to the global sequence surveys (see below) and still in common use today, is the use of qPCR assays to quantify the *nifH* gene for diazotrophs. The *nifH* abundances of the various DDA symbionts (het-1, het-2, and het-3) co-occur and can even outnumber UCYN-A *nifH* copy abundances, however this is rarely highlighted.

We reviewed 46 qPCR *nifH* surveys (reviewed in Farnelid et al., 2016; Stenegren et al., 2017), and found that in <30% of the studies (13 of 46 studies) all three DDA symbionts were quantified, compared to UCYN-A (96%, 44 of 46 studies) and the colonial diazotroph *Trichodesmium* (76%, 35 of 46 studies). A higher percent (ca 50%, 24 of 46 studies) quantified het-1 and to a lesser extent het-2 (ca 40%, 19 of 46 studies). Additionally, often only small subsets of the total samples were assayed. For example, 19 out of 154 samples (12%) were assayed for het-1 and het-2 in an area of the North Atlantic known for large-scale blooms of DDAs (Goebel et al., 2010). Hence it seems there is an "un-quantified disparity" against the DDA symbionts that is especially disconcerting in the context that often the global significance of an organism is associated with its abundance and distribution. Moreover, other means to collect the DDAs, e.g., gravity filtration from niskin bottles, net tows, perhaps are still a better means to quantify large and rare symbiosis.

Recent advances in sequencing platforms combined with large scale oceanic sampling surveys [e.g., global ocean survey (GOS), TARA, Malaspina] have been steadfast at revealing the rich diversity of planktonic populations in the global ocean without the need of separating populations (Caporaso et al., 2011, 2012; Werner et al., 2012; Bokulich et al., 2013). The UCYN-A based symbioses (both DNA and RNA based samples) have been reported in the TARA project and from a few select stations (Cabello et al., 2016; Cornejo-Castillo et al., 2016), while the DDAs remain un-reported. Using several genetic markers derived from the DDA symbionts (16S rRNA, *nifH*) and hosts (*rbcL*, 18S rRNA), we performed blast searches in the KEGG MGENES (www.genome.jp/mgenes/) and in the IMNGS database (https://www.imngs.org/), with a 95% identity cut off. Our searches did not generate a single hit, which was surprising, since DDAs are widespread and often bloom formers. The closest hit were at the 90% sequence identity with low coverage.

A closer look at the sampling methods employed in the global surveys identifies the use of pre-filtration steps (e.g., varying from 20 to 200 μm). The pre-filtration is necessary to avoid inclusion of larger eukaryotes (e.g., gelatinous zooplankton, larvae and eggs), which would dominate the nucleic acid composition and hence the sequence libraries. The hosts of the DDAs are large (>20 μm) and often form colonies or chains (>200 μm), thus pre-filtration would systematically exclude DDAs. The prymnesiophyte-UCYN-A symbioses are small in comparison; both hosts and symbionts are <10 μm. In fact a closer examination at the size fractionation experiments reported in Tripp et al. (2010) (Figure 4, Tripp et al., 2010), demonstrates

**TABLE 1 |** Summary of characters for DDA and Prymnesiophyte-UCYN-A symbioses are shown and separated by symbiont identity and common abbreviation for each symbiont is given in parenthesis (e.g. het-1)[†].

| R. intracellularis (het-1) | R. intracellularis (het-2) | C. rhizosoleniae (het-3) | Character | A. thalassa (UCYN-A1) | A. thalassa (UCYN-A2) |
|---|---|---|---|---|---|
| √R. clevei R. styliformis | H. hauckii (HH) H. membranaceus (HM) | C. compressus Bacteriastrum | **known host(s)** | B. bigelowii relative | B. bigelowii |
| 80–250 µm | 12–35 µm 30–70 µm | 7–40 µm | **host size*** | 1–3 µm | 4–5 µm; 7–10 µm |
| chains and singlet | chains and singlet | chains | **host life history** | single | single |
| endobiont | Presumed endobiont | external | **symbiont location** | endobiont | endobiont, symbiosome |
| 1–32 | 1–2 | 1–4 | **# symbionts/ host** | 1 | 3–10 |
| | 5–6 µm; trichome length varies | | **symbiont cell dia.** | 1 µm | >1 µm |
| 5.4 | 3.24 (RintHH) 2.21 (RintHM) | 5.97 | **Symbiont genome size (Mbp)** | 1.44 | 1.48 |
| 39 | 34 (RintHH) 34 (RintHM) | 39 | **GC content (%)** | 31 | 31 |
| Draft (857) | RintHH draft (90) RintHM draft (941) | Draft (620) | **Genome status (contigs)** | Complete (1) | Draft (52) |
| 65 | 56 (RintHH) | 77 | **Coding density (%)** | 81 | 79 |
| NH$_4^+$ transporter; NO$_3$ transporter and reductase; NO$_2$ reductase; Urea transporter and urease; GS inactivating factor | NH$_4^+$ transporter; NO$_3$ transporter and reductase; NO$_2$ reductase; Urea transporter and urease; GS inactivating factor; GOGAT | Urea transporter and urease; GS inactivating factor | **reported genome deletions** | PS-II, partial or incomplete: TCA, Calvin and urea cycles; amino acids and purine synthesis | |
| provide N | provide N | provide N | **Symbiont function** | provide N | provide N |
| unknown | unknown | unknown | **Known host function** | provide C | provide C |
| 37.0 | 29.9 | 41.5 | **# Transporters/Mbp[^]** | 41.0 | 41.2 |
| 3.1 | 6.8 | 4.5 | **# Genes for N$_2$ fix/Mbp[^]** | 11.8 | 9.7 |
| 27.2 | 29.6 | 22.6 | **# Genes for carbon/Mbp**** | 19.4 | 20.1 |

[†] *References: Villareal, 1992; Foster and Zehr, 2006; Foster and O'Mullan, 2008; Zehr et al., 2008; Tripp et al., 2010; Thompson et al., 2012, 2014; Hagino et al., 2013; Hilton et al., 2013; Krupke et al., 2013; Bombar et al., 2014; Hilton, 2014; Cabello et al., 2016; Cornejo-Castillo et al., 2016; Martínez-Pérez et al., 2016.* √ *up to 13 species are reported for Rhizosolenia-R. intracellularis symbioses hence we report the two most common.*
*\*Diatom cell size are reported as apical axis length.*
*\*\*, #genes for carbon includes genes for photosystems, light-harvesting pigments and CO$_2$ uptake.*
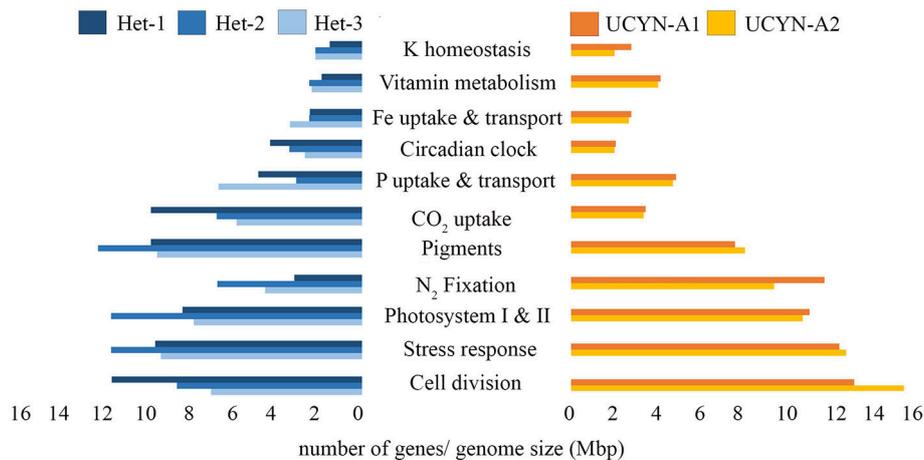*[^]Mbp × 10$^6$.*

that even though the UCYN-A hosts are small (e.g., 4–5 or 7–10 µm), the UCYN-A *nifH* genes were still detected in the >10 and >25 µm size fractions. Detecting UCYN-A on the larger size fractions suggests that even the UCYN-A symbioses are likely under-sampled in the global surveys. Future sampling campaigns (and interpretations of results) should consider and modify sampling to avoid biases against larger, symbiotic, and/or chains of cells.

## SINGLE CELL N$_2$ FIXATION RATES AND SCALING UP TO THE WATER COLUMN

Measurements of N$_2$ fixation and transfer of fixed N from symbiont to host has been made in both symbioses using $^{15}$N$_2$ and nano-meter scale secondary ion-mass spectrometry (nanoSIMS) measurements (Foster et al., 2011; Thompson et al., 2012; Krupke et al., 2013, 2014; Martínez-Pérez et al., 2016).

Since C fixation has only been reported in the prymnesiophyte-UCYN-A symbioses, we limit the comparison to the N$_2$ fixation measures. For the prymnesiophyte-UCYN-A symbiosis, a halogenated *in situ* hybridization (HISH) step is required prior to the SIMS measurements for identification of the UCYN-A symbiont, and in a more recent study, an additional probe specific for the UCYN-A hosts were used, hence a dual labeling (Martínez-Pérez et al., 2016).

A closer look at the single N$_2$ fixation activity by DDA and UCYN-A symbionts highlights a surprising similarity. For example, N$_2$ fixation for the *Hemiaulus-Richelia* DDA range between 1.15 and 50.4 fmol N cell$^{-1}$ h$^{-1}$ (Foster et al., 2011), while recently N$_2$ fixation by UCYN-A1 and A2 was 0.5 and 9.16 fmol N cell$^{-1}$ h$^{-1}$, respectively (Martínez-Pérez et al., 2016). Both UCYN-A and the DDA symbionts function as N sources for their respective hosts. Given that *Hemiaulus* spp. hosts are larger (12–70 µm) than the prymnesiophyte hosts (1–10 µm), one expected a higher N$_2$ fixation in the DDAs to support the higher N demand

**FIGURE 1 |** Subsystem (genes with similar function) comparison between DDA and UCYN-A symbionts. Each bar indicates the number of genes in a particular subsystem normalized against genome size (Mbp × $10^6$) for het-1 (dark blue), het-2 (lighter-blue), het-3 (lightest blue), UCYN-A1 (dark orange) and UCYN-A2 (light orange).

of the larger host diatoms. However, other factors influence $N_2$ fixation, e.g., host growth, light/energy availability, temperature, and iron (Villareal, 1989; Kustka et al., 2002; Berman-Frank et al., 2007).

The recent rates reported for UCYN-A1 were also 3–5 orders of magnitude higher than the rates reported earlier for UCYN-A1 (e.g., $6.35 \times 10^{-5} - 0.013 \times 10^{-2}$ fmol N $cell^{-1}$ $h^{-1}$), despite similar measured enrichment values (e.g., 0.3744-0.6194 atom%$^{15}$N) (Krupke et al., 2013). Inherent in using the HISH-SIMS is the potential to reduce the $^{15}$N fraction in the cell (e.g., up to 60%) (Musat et al., 2014). Hence it was somewhat surprising that in the recent work on UCYN-A1 that utilized a double CARD-FISH measured such a higher enrichment in $^{15}$N than the earlier works that used only a single probe. The effect of double CARD-FISH on enrichment has not been tested.

One of the useful applications for the single cell rates is the ability to estimate the contribution of a particular population to bulk water activity. These types of estimates have been done for several of the open ocean diazotrophs (Foster et al., 2011, 2013; Krupke et al., 2015; Martínez-Pérez et al., 2016). For example, an estimate of 0.19 and 0.62 Tmol N $yr^{-1}$ is provided by DDAs in the Atlantic and Pacific oceans, respectively, which is similar to 0.36–0.71 Tmol N $yr^{-1}$ estimated for *Trichodesmium* spp. (Capone et al., 2008; Foster et al., 2011). Recently, up to 30% of $N_2$ fixation in the surface tropical N. Atlantic was attributed to UCYN-A symbioses, while the contribution by *Trichodesmium* in the region was far less (11%) (Martínez-Pérez et al., 2016). Hence, single cell activity measures have demonstrated that both symbioses should be included in global N models.

## CONCLUSIONS

In summary, even though DDAs are one of the earliest recorded symbioses and are frequently observed at high bloom densities worldwide with high measured $N_2$ fixation rates leading to

high export fluxes (Villareal, 1994; Carpenter et al., 1999; Foster et al., 2007, 2011; Subramaniam et al., 2008; Karl et al., 2012), the DDAs are often overlooked. The genomes for the UCYN-A lineages and one of the het symbionts (RintHH01) are uniquely streamlined and similarly optimized for a symbiotic life history. An open and interesting question is whether a similar degree of modification occurs in the respective host genomes. Both symbioses have invested in $N_2$ fixation in both genome content and cellular activity. Moreover, recent evidence suggests that both symbioses can have equally high $N_2$ fixation rates and should be recognized as major players in the N (and C) cycle. DDAs are methodologically excluded and less often assayed in qPCR *nifH* surveys leading to under-detection, while UCYN-A are often the most frequently assayed, and thus unavoidably the most detected diazotroph. In a time of incredible resources and instrumentation for detection, one must consider the "un-quantified disparity," especially if sequence abundance and information continues to be used as a comparative indices and to predict biogeochemical significance.

## AUTHOR CONTRIBUTIONS

All authors contributed to the manuscript. AC, MS, and RF compiled the content of **Table 1** and MS compiled the data and generated **Figure 1**.

# REFERENCES

Berman-Frank, I., Quigg, A., Finkel, Z. V., Irwin, A. J., and Haramaty, L. (2007). Nitrogen-fixation strategies and Fe requirements in cyanobacteria. *Limnol. Oceanogr.* 52, 2260–2269. doi: 10.4319/lo.2007.52.5.2260

Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, I. J., Knight, R., et al. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Methods* 10, 57–59. doi: 10.1038/nmeth.2276

Bombar, D., Heller, P., Sanchez-Baracaldo, P., Carter, B. J., and Zehr, J. P. (2014). Comparative genomics reveals surprising divergence of two closely related strains of uncultivated UCYN-A cyanobacteria. *ISME J.* 8, 2530–2542. doi: 10.1038/ismej.2014.167

Bothe, H., Tripp, H. J., and Zehr, J. P. (2010). Unicellular cyanobacteria with a new mode of life: the lack of photosynthetic oxygen evolution allows nitrogen fixation to proceed. *Arch. Microbiol.* 192, 783–790. doi: 10.1007/s00203-010-0621-5

Cabello, A. M., Cornejo-Castillo, F. M., Raho, N., Blasco, D., Vidal, M., Audic, S., et al. (2016). Global distribution and vertical patterns of a prymnesiophyte–cyanobacteria obligate symbiosis. *ISME J.* 10, 693–706. doi: 10.1038/ismej.2015.147

Capone, D. G., Bronk, D. A., Mulholland, M. R., and Carpenter, E. J. (eds.). (2008). *Nitrogen in the Marine Environment.* London: Academic Press, 1758.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624. doi: 10.1038/ismej.2012.8

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-lyons, D., Lozupone, C. A., Turnbaugh, P. J., et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad.Sci. U.S.A.* 108, 4516–4522. doi: 10.1073/pnas.1000080107

Carpenter, E. J., and Janson, S. (2000). Intracellular cyanobacterial symbionts in the marine diatom *Climacodium frauenfeldianum* (Bacillariophyceae). *J. Phycol.* 36, 540–544. doi: 10.1046/j.1529-8817.2000.99163.x

Carpenter, E. J., Montoya, J. P., Burns, J., Mulholland, M. R., Subramaniam, A., and Capone, D. G. (1999). Extensive bloom of a $N_2$-fixing diatom/cyanobacterial association in the tropical Atlantic Ocean. *Mar. Ecol. Prog. Ser.* 185, 273–283. doi: 10.3354/meps185273

Church, M. J., Short, C. M., Jenkins, B. D., Karl, D. M., and Zehr, J. P. (2005). Temporal patterns of nitrogenase gene (*nifH*) expression in the oligotrophic North Pacific ocean. *Appl. Environ. Microbiol.* 71, 5362–5370. doi: 10.1128/AEM.71.9.5362-5370.2005

Cornejo-Castillo, F. M., Cabello, A. M., Salazar, G., Sánchez-Baracaldo, P., Lima-mendez, G., Hingamp, P., et al. (2016). Cyanobacterial symbionts diverged in the late Cretaceous towards lineage-specific nitrogen fixation factories in single-celled phytoplankton. *Nat. Commun.* 7, 1–9. doi: 10.1038/ncomms11071

Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J., and Smith, A. G. (2005). Algae acquire vitamin $B_{12}$ through a symbiotic relationship with bacteria. *Nature* 438, 90–93. doi: 10.1038/nature04056

Croft, M. T., Warren, M. J., and Smith, A. G. (2006). Algae need their vitamins. *Eukaryotic Cell* 5, 1175–1183. doi: 10.1128/EC.00097-06

Delong, E. F. (2010). Interesting things come in small packages. *Genome Biol.* 11:118. doi: 10.1186/gb-2010-11-5-118

Farnelid, H., Turk-Kubo, K., Muñoz-Marín, M., and Zehr, J. (2016). New insights into the ecology of the globally significant uncultured nitrogen-fixing symbiont UCYN-A. *Aquat. Microb. Ecol.* 77, 125–138. doi: 10.3354/ame01794

Foster, R. A., Goebel, N. L., and Zehr, J. P. (2010). Isolation of *Calothrix rhizosoleniae* (cyanobacteria) strain SC01 from *Chaetoceros* (Bacillariophyta) spp. diatoms of the subtropical North Pacific Ocean. *J. Phycol.* 46, 1028–1037. doi: 10.1111/j.1529-8817.2010.00885.x

Foster, R. A., Kuypers, M. M. M., Vagner, T., Paerl, R. W., Musat, N., and Zehr, J. P. (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., and O'Mullan, G. D. (2008). "Nitrogen-fixing and nitrifying symbioses in the marine environment," in *Nitrogen in the Marine Environment*, eds D. G. Capone, D. A. Bronk, M. R. Mulholland, and E. J. Carpenter (London: Academic Press), 1197–1218.

Foster, R. A., Subramaniam, A., Mahaffey, C., Carpenter, E. J., Capone, D. G., and Zehr, J. P. (2007). Influence of the Amazon River plume on distributions of free-living and symbiotic cyanobacteria in the western tropical north Atlantic Ocean. *Limnol. Oceanogr.* 52, 517–532. doi: 10.4319/lo.2007.52.2.0517

Foster, R. A., Sztejrenszus, S., and Kuypers, M. M. (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

Foster, R. A., and Zehr, J. P. (2006). Characterization of diatom-cyanobacteria symbioses on the basis of *nifH*, *hetR* and 16S rRNA sequences. *Environ. Microbiol.* 8, 1913–1925. doi: 10.1111/j.1462-2920.2006.01068.x

Goebel, N. L., Turk, K. A., Achilles, K. M., Paerl, R., Hewson, I., Morrison, A. E., et al. (2010). Abundance and distribution of major groups of diazotrophic cyanobacteria and their potential contribution to $N_2$ fixation in the tropical Atlantic Ocean. *Environ. Microbiol.* 12, 3272–3289. doi: 10.1111/j.1462-2920.2010.02303.x

Gómez, F., Furuya, K., and Takeda, S. (2005). Distribution of the cyanobacterium *Richelia intracellularis* as an epiphyte of the diatom *Chaetoceros compressus* in the western Pacific Ocean. *J. Plankton Res.* 27, 323–330. doi: 10.1093/plankt/fbi007

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

Hilton, J. A. (2014). *Ecology and Evolution of Diatom-Associated Cyanobacteria Through Genetic Analyses.* Dissertation's thesis. Santa Cruz, CA: University of California.

Hilton, J. A., Foster, R. A., Tripp, H. J., Carter, B. J., Zehr, J. P., and Villareal, T. A. (2013). Genomic deletions disrupt nitrogen metabolism pathways of a cyanobacterial diatom symbiont. *Nat. Commun.* 4, 1–7. doi: 10.1038/ncomms2748

Janson, S., Woulters, J., Bergman, B., and Carpenter, E. (1999). Host specificity in the *Richelia*-diatom symbiosis revealed by *hetR* gene sequence analysis. *Environ. Microbiol.* 1, 431–438. doi: 10.1046/j.1462-2920.1999.00053.x

Karl, D. M., Church, M. J., Dore, J. E., Letelier, R. M., and Mahaffey, C. (2012). Predictable and efficient carbon sequestration in the North Pacific Ocean supported by symbiotic nitrogen fixation. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1842–1849. doi: 10.1073/pnas.1120312109

Krupke, A., Lavik, G., Halm, H., Fuchs, B. M., Amann, R. I., and Kuypers, M. M. (2014). Distribution of a consortium between unicellular algae and the $N_2$ fixing cyanobacterium UCYN-A in the North Atlantic Ocean. *Environ. Microbiol.* 16, 3153–3167. doi: 10.1111/1462-2920.12431

Krupke, A., Mohr, W., Laroche, J., Fuchs, B. M., Amann, R. I., and Kuypers, M. M. (2015). The effect of nutrients on carbon and nitrogen fixation by the UCYN-A-haptophyte symbiosis. *ISME J.* 9, 1635–1647. doi: 10.1038/ismej.2014.253

Krupke, A., Musat, N., LaRoche, J., Mohr, W., Fuchs, B. M., Amann, R. I., et al. (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

Kustka, A., Carpenter, E. J., and Sañudo-Wilhelmy, S. A. (2002). Iron and marine nitrogen fixation: progress and future directions. *Res. Microbiol.* 153, 255–262. doi: 10.1016/S0923-2508(02)01325-6

Lemmermann, E. (1905). Die algenflora der sandwich-inseln. ergebnisse einer reise nach dem Pacific. H. Schauinsland 1896/97. *Bot. Jahrbuch Syst. Pflanzengesch. Pflanzengeogr.* 34, 607–663.

Martínez-Pérez, C., Mohr, W., Löscher, C. R., Dekaezemacker, J., Littmann, S., Yilmaz, P., et al. (2016). The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle. *Nat. Microbiol.* 1:16163. doi: 10.1038/nmicrobiol.2016.163

Marumo, R., and Asaoka, O. (1974). Distribution of pelagic blue-green algae in the North Pacific Ocean. *J. Oceanogr. Soc. Japan* 30, 77–85. doi: 10.1007/BF02112896

Musat, N., Stryhanyuk, H., Bombach, P., Adrian, L., Audinot, J. N., and Richnow, H. H. (2014). The effect of FISH and CARD-FISH on the isotopic composition of $^{13}$C- and $^{15}$N-labeled *Pseudomonas putida* cells measured by nanoSIMS. *Syst. Appl. Microbiol.* 37, 267–276. doi: 10.1016/j.syapm.2014.02.002

Ostenfeld, C. H., and Schmidt, J. (1901). *Plankton fra det Rode Hav og Adenbugten (Plankton from the Red Sea and the Gulf of Aden).* Copenhagen.

Rasmussen, U. (2002). "Cyanobacterial diversity and specificity in plant symbioses," in *Cyanobacteria in Symbiosis*, eds A. N. Rai, B. Bergman, and U. Rasmussen (Dordrecht: Kluwer Academic Press), 313–328.

Sournia, A. (1970). Les cyanophycees dans le plancton marin. *Ann. Biol.* 9, 63–76.

Stenegren, M., Caputo, A., Berg, C., Bonnet, S., and Foster, R. A. (2017). Distribution and drivers of 5 symbiotic and free-living diazotrophic cyanobacteria in the Western Tropical South Pacific. *Biogeosci. Discuss.* 2017, 1–47, doi: 10.5194/bg-2017-63

Subramaniam, A., Yager, P. L., Carpenter, E. J., Mahaffey, C., Björkman, K. M., Cooley, S., et al. (2008). Amazon river enhances diazotrophy and carbon sequestration in the tropical North Atlantic ocean. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10460–10465. doi: 10.1073/pnas.0710279105

Thompson, A., Carter, B. J., Turk-Kubo, K., Malfatti, F., Azam, F., and Zehr, J. P. (2014). Genetic diversity of the unicellular nitrogen-fixing cyanobacteria UCYN-A and its prymnesiophyte host. *Environ. Microbiol.* 16, 3238–3249. doi: 10.1111/1462-2920.12490

Thompson, A. W., Foster, R. A., Krupke, A., Carter, B. J., Musat, N., Vaulot, D., et al. (2012). Unicellular cyanobacterium symbiotic with a single-celled eukaryotic alga. *Science* 337, 1546–1550. doi: 10.1126/science.1222700

Thompson, A. W., and Zehr, J. P. (2013). Cellular interactions: lessons from the nitrogen-fixing cyanobacteria. *J. Phycol.* 49, 1024–1035. doi: 10.1111/jpy.12117

Tripp, H. J., Bench, S. R., Turk, K. A., Foster, R. A., Desany, B. A., Niazi, F., et al. (2010). Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium. *Nature* 464, 90–94. doi: 10.1038/nature08786

Turk-Kubo, K. A., Farnelid, H. M., Shilova, I. N., Henke, B., and Zehr, J. P. (2017). Distinct ecological niches of marine symbiotic N$_2$-fixing cyanobacterium Candidatus *Atelocyanobacterium thalassa* sublineages. *J. Phycol.* 461, 451–461. doi: 10.1111/jpy.12505

Villareal, T. A. (1989). Division cycles in the nitrogen-fixing *Rhizosolenia* (Bacillariophyceae)-*Richelia* (Nostocaceae) symbiosis. *Br. J. Phycol.* 24, 357–365. doi: 10.1080/00071618900650371

Villareal, T. A. (1990). Laboratory culture and preliminary characterization of the nitrogen-fixing *Rhizosolenia-Richelia* symbiosis. *Mar. Ecol.* 11, 117–132. doi: 10.1111/j.1439-0485.1990.tb00233.x

Villareal, T. A. (1992). "Marine nitrogen-fixing diatom-Cyanobacteria symbioses," in *Marine Pelagic Cyanobacteria: Trichodesmium and other Diazotrophs*, ed E. J. Carpenter (Dordrecht: Springer), 163–175.

Villareal, T. A. (1994). Widespread occurrence of the *Hemiaulus*-cyanobacterial symbiosis in the Southwest North Atlantic Ocean. *Bull. Mar. Sci.* 54, 1–7.

Werner, J. J., Zhou, D., Caporaso, J. G., Knight, R., and Angenent, L. T. (2012). Comparison of Illumina paired-end and single-direction sequencing for microbial 16S rRNA gene amplicon surveys. *ISME J.* 6, 1273–1276. doi: 10.1038/ismej.2011.186

White, A. E., Prahl, F. G., Letelier, R. M., and Popp, B. N. (2007). Summer surface waters in the Gulf of California: prime habitat for biological N$_2$-fixation. *Global Biogeochem. Cycles* 21, 1–11. doi: 10.1029/2006GB002779

Zehr, J. P. (2015). How single cells work together. *Science* 349, 1163–1164. doi: 10.1126/science.aac9752

Zehr, J. P., Bench, S. R., Carter, B. J., Hewson, I., Niazi, F., Shi, T., et al. (2008). Globally distributed uncultivated oceanic N$_2$-fixing cyanobacteria lack oxygenic photosystem II. *Science* 322, 1110–1112. doi: 10.1126/science.1165340

Zehr, J. P., Mellon, M. T., Zani, S., and York, N. (1998). New Nitrogen-fixing microorganisms detected in oligotrophic oceans by amplification of nitrogenase (*nifH*) genes. *Appl. Environ. Microbiol.* 64, 3444–3450.

Zehr, J. P., Shilova, I. N., Farnelid, H. M., Muñoz-marín, M. C., and Turk-Kubo, K. A. (2016). Unusual marine unicellular symbiosis with the nitrogen-fixing cyanobacterium UCYN-A. *Nat. Microbiol.* 2:16214. doi: 10.1038/nmicrobiol.2016.214

Zehr, J., Waterbury, J., and Turner, P. (2001). Unicellular cyanobacteria fix N$_2$ in the subtropical North Pacific Ocean. *Nature* 715, 25–28. doi: 10.1038/35088063

# SUPPLEMENTARY METHOD FOR CONSTRUCTING THE SUBSYSTEMS COMPARISON (FIGURE 1)

Genomes were downloaded from the NCBI database (RintRC01 accession: PRJEB5207, RintHH01 accession: PRJEA104979, RINTHM01 accession: PRJEA104981, CalSC01 accession: PRJEB19506, UCYN-A1 ALOHA accession: PRJNA30917, UCYN-A2 SIO64986 accession: PRJNA256120) and subsequently submitted to RAST (Rapid Annotation using Subsystem Technology) for annotation. The annotated genomes where browsed for genetic pathways of interest by using key word searches. Number of genes for a particular subsystem were normalized to genome size.