



REVIEW

New insights into the ecology of the globally significant uncultured nitrogen-fixing symbiont UCYN-A

Hanna Farnelid^{1,2,*}, Kendra Turk-Kubo¹, María del Carmen Muñoz-Marín^{1,3}, Jonathan P. Zehr¹

¹Ocean Sciences Department, University of California, Santa Cruz, CA 95064, USA

²Centre for Ecology and Evolution in Microbial Model Systems (EEMiS), Linnaeus University, 392 34 Kalmar, Sweden

³Departamento de Bioquímica y Biología Molecular, Edificio Severo Ochoa, planta 1, Universidad de Córdoba, 14071-Córdoba, Spain

ABSTRACT: Cyanobacterial nitrogen-fixers (diazotrophs) play a key role in biogeochemical cycling of carbon and nitrogen in the ocean. In recent years, the unusual symbiotic diazotrophic cyanobacterium *Atelocyanobacterium thalassa* (UCYN-A) has been recognized as one of the major diazotrophs in the tropical and subtropical oceans. In this review, we summarize what is currently known about the geographic distribution of UCYN-A, as well as the environmental factors that govern its distribution. In addition, by compiling UCYN-A *nifH* sequences from the GenBank no. database as well as those from *nifH* gene amplicon next generation sequencing studies, we present an in-depth analysis of the distribution of defined UCYN-A sublineages (UCYN-A1, UCYN-A2 and UCYN-A3) and identify a novel sublineage, UCYN-A4, which may be significant in some environments. Each UCYN-A sublineage exhibited a remarkable global distribution pattern and several UCYN-A sublineages frequently co-occurred within the same sample, suggesting that if they represent different ecotypes they have overlapping niches. Recently, single cell visualization techniques using specific probes targeting UCYN-A1 and UCYN-A2 and their respective associated eukaryotic partner cells showed that the size of the consortia and the number of UCYN-A cells differed between these 2 sublineages. Combined, the results highlight that UCYN-A sublineages likely have different physiological requirements, which need to be accounted for in future studies. Furthermore, based on our increasing knowledge of the diversity of the UCYN-A lineage, we discuss some of the limitations of currently used cultivation-independent molecular techniques for the identification and quantification of UCYN-A.

KEY WORDS: UCYN-A · Symbiosis · *nifH* · qPCR · Next generation amplicon sequencing · CARD-FISH · *Braarudosphaera bigelowii*

INTRODUCTION

In the oceans, the availability of fixed nitrogen (N) is an important factor that limits biological productivity (Ryther & Dunstan 1971, Moore et al. 2013). Nitrogen (N₂) fixation is an enzyme-mediated process that is

found among diverse lineages of *Bacteria* and *Archaea*, providing biologically available ammonium (NH₄⁺) from dinitrogen gas. N₂-fixers (diazotrophs) can support up to 50 % of the new production (Karl et al. 1997) and thereby play a central role in the fixation of carbon (CO₂) and its export out of the photic zone.

*Corresponding author: hannafarnelid@gmail.com

For many years, the paradigm was that N_2 fixation in the oceans was due primarily to *Trichodesmium* (Capone et al. 1997, 2005) and diatom symbionts (e.g. Villareal & Carpenter 1989, Villareal 1990, 1991, 1994, Subramaniam et al. 2008) and was restricted to warm, oligotrophic waters in tropical seas. However, using molecular approaches targeting the *nifH* gene, which encodes the iron protein of the nitrogenase enzyme, additional marine diazotrophs including unicellular cyanobacteria and heterotrophic bacteria were discovered in the 1990s (Zehr et al. 1998, 2001). Today it is recognized that unicellular cyanobacteria, including the uncultivated *Atelocyanobacterium thalassa* (UCYN-A) and *Crocospaera watsonii* (UCYN-B), contribute significantly to global N_2 fixation and may equal that of *Trichodesmium* (Montoya et al. 2004, Moisander et al. 2010). Early studies indicated that the UCYN-A group was of particular interest, as a fragment of its *nifH* gene was recovered from multiple ocean basins including the North Pacific and North Atlantic (Zehr et al. 1998, 2001, Falcón et al. 2004, Church et al. 2005a, Langlois et al. 2005, Fong et al. 2008). UCYN-A *nifH* genes and *nifH* gene transcripts, a proxy for active N_2 fixation, were found at higher latitudes and deeper in the water column than *Trichodesmium* (Moisander et al. 2010). Recent studies also indicated that the distribution of UCYN-A goes beyond that of the oligotrophic oceans, and thereby further extends the known geographic range of N_2 fixation. Since the discovery of UCYN-A, knowledge of the global distribution and significance of UCYN-A has continued to expand but there remain large unknowns about its ecology and physiology.

Observations of the transcription pattern of the UCYN-A *nifH* gene (Church et al. 2005b, Zehr et al. 2007), with expression during the day, hinted at a different physiology compared to other unicellular N_2 -fixing cyanobacteria. Subsequently, sequencing of the UCYN-A genome revealed an extremely unusual phototrophic microorganism that lacked the genes for oxygenic photosynthesis, CO_2 fixation and the entire tricarboxylic acid (TCA) cycle, as well as other metabolic pathways (Zehr et al. 2008, Tripp et al. 2010). The significant genome reduction found in these studies led to the hypothesis that UCYN-A must live in symbiosis. Shortly after, a small prymnesiophyte (haptophyte), related to *Braarudosphaera bigelowii*, was identified as the symbiotic partner cell (Thompson et al. 2012, Hagino et al. 2013). This relationship is the only known example of a symbiosis between a diazotroph and a haptophyte and resembles that of the single cell associations that may have led to the evolution of plastids (Zehr 2015). The basis

of this symbiosis is the exchange of fixed carbon (C; from the phototrophic eukaryote) and fixed N_2 (from the diazotrophic UCYN-A) (Thompson et al. 2012, Krupke et al. 2013, 2015). Further, the reduced genome of UCYN-A and tight coupling of N and C transfers between UCYN-A and its eukaryotic partner suggests a mutually beneficial and obligate relationship (Thompson et al. 2012, Krupke et al. 2014, 2015). Recent studies suggest that UCYN-A and its eukaryotic partner cell have likely co-evolved (Bombar et al. 2014, Cornejo-Castillo et al. 2016), yet the nature of the physical association between the cells is currently not well characterized (Zehr 2015).

Amplicon sequencing of the *nifH* gene, both using clone and next generation sequencing libraries and targeted quantitative PCR (qPCR) assays have continued to expand our knowledge of the distribution patterns of UCYN-A. Recently, genetic diversity among UCYN-A *nifH* sequences was reported and 3 sublineages were defined based on the sequences available in GenBank at the time (Thompson et al. 2014). Currently, there are 2 genome sequences from 2 different UCYN-A sublineages (Zehr et al. 2008, Tripp et al. 2010, Bombar et al. 2014). The organisms, UCYN-A1 and UCYN-A2, have similar gene content but also show remarkable differences (Bombar et al. 2014). The eukaryotic partner cell for UCYN-A2 was identified as being closely related, but not identical to the prymnesiophyte partner of UCYN-A1 (based on 18S rRNA gene sequence similarity; Thompson et al. 2012, 2014). Sequencing of these genomes and identification of the associated eukaryotic cells have been critical milestones in the study of UCYN-A, without which the recent advances in knowledge of this unusual symbiosis would have been impossible. For example, whole genome expression patterns of UCYN-A were recently studied as part of a large metatranscriptomic effort during the Tara-Oceans Expedition, and specific molecular probes have been developed which have enabled, for the first time, the visualization of UCYN-A1 and UCYN-A2 and their respective partner cells (Cabello et al. 2016, Cornejo-Castillo et al. 2016).

In this review, we discuss current knowledge about UCYN-A with respect to recent methodological milestones. To identify global distribution patterns of UCYN-A, we synthesized data from all studies reporting UCYN-A abundances using quantitative techniques. In addition, we compiled all available UCYN-A *nifH* gene sequences, including GenBank nr database sequences as well as those reported from next generation sequencing studies. With this significant increase in numbers of sequences, a novel

UCYN-A sublineage was identified and patterns in the global distribution of the UCYN-A sublineages begin to emerge. We also discuss recent methodological advancements in visualization techniques, which have provided valuable insights into the morphology and physiology of UCYN-A and the eukaryotic partner cell.

ESTIMATION OF UCYN-A N_2 FIXATION RATES, GLOBAL DISTRIBUTION AND ABUNDANCES

Since UCYN-A has not yet been cultured and it frequently co-occurs with other diazotrophs, studying specific N_2 fixation rates by UCYN-A is difficult. By size-partitioning the bulk seawater community into larger and smaller size fractions, it has been found that the smaller size fraction can at times dominate water column N_2 fixation rates (Montoya et al. 2004, Garcia et al. 2007, Biegala & Raimbault 2008, Bonnet et al. 2009, Church et al. 2009, Benavides et al. 2011). The importance of unicellular cyanobacteria has been inferred from regions where N_2 fixation was measured in small size-fractions along with the presence and, in some cases, high abundances of *nifH* genes from unicellular cyanobacteria (including, for example, UCYN-A and UCYN-B). The small size-fraction (generally defined as 0.2 to 10 μm in most of these studies) is frequently dominated by UCYN-A, and significant N_2 fixation rates have therefore been attributed to UCYN-A at, for example, Stn ALOHA in the North Pacific (Montoya et al. 2004, Church et al. 2009). Interestingly, UCYN-A has not been detected exclusively in the small size-fraction (Benavides et al. 2016, Cornejo-Castillo et al. 2016), suggesting there is a size distribution within the UCYN-A population. Consequently, size-fractionation may exclude UCYN-A if they are associated with a larger eukaryotic cell, and thereby N_2 fixation rates by UCYN-A could have been underestimated.

With recent single cell methodological developments, the metabolic activity of specific UCYN-A cells can be studied. Using halogenated *in situ* hybridization (HISH) or catalyzed reporter deposition-fluorescence *in situ* hybridization (CARD-FISH) in combination with stable isotope labeling experiments and high-resolution nanometer scale secondary ion mass spectrometry (nanoSIMS), it was demonstrated that a large fraction of the fixed N was transferred to the associated eukaryotic partner, while a smaller fraction of fixed C was transferred between the eukaryotic cell and UCYN-A (Thompson et al. 2012, Krupke et al. 2015). However, the N_2 fixation rates

measured for single UCYN-A cells (Thompson et al. 2012, Krupke et al. 2013, 2015) were lower than expected based on modeled rates determined using growth and biomass of UCYN-A (Goebel et al. 2008). The cellular ^{13}C and ^{15}N enrichments may be underestimated due to the decrease of signal in association with CARD-FISH analysis, a necessary technique to identify UCYN-A cells prior to nanoSIMS analysis (Musat et al. 2014). Although nanoSIMS studies targeting UCYN-A are thus far limited in both numbers and geographic coverage, it will likely become an important tool for determining the range of cell-specific N_2 fixation rates as well as physiological responses of UCYN-A and its eukaryotic partner cell to perturbations, which will result in more accurate estimates of the contribution to marine N_2 fixation.

Since the first discovery of unicellular cyanobacterial diazotrophs, *nifH* gene PCR amplicon libraries and species-specific *nifH* gene qPCR assays have been used in numerous studies at various locations to detect the presence of, and to quantify, genes and/or transcripts. Wide distribution patterns, high abundances and transcriptional activity of UCYN-A *nifH* genes have been reported from diverse locations, further supporting the significance of UCYN-A to marine N_2 fixation. In the tropical oligotrophic ocean, abundances of close to 10^7 *nifH* gene copies l^{-1} have been reported (Fig. 1, Table S1 in Supplement 1 at www.int-res.com/articles/suppl/a077p125_supp.pdf) which corresponds to high cellular abundances of UCYN-A. Compared to *Trichodesmium*, which tends to be more abundant in the upper water column (Luo et al. 2012), UCYN-A were found to occupy greater depths (down to ~150 m) and were present at higher latitudes, extending the previously known N_2 fixation domains (Moisander et al. 2010). Interestingly, recent studies report UCYN-A presence at diverse marine locations and environmental conditions including coastal, high latitude, N-replete and hypersaline environments, as well as in marine sediments and within copepod guts (Brown & Jenkins 2014, Bentzon-Tilia et al. 2015, Messer et al. 2015, Scavotto et al. 2015). Using CARD-FISH, Cabello et al. (2016) detected UCYN-A in surface waters spanning the circumglobal MALASPINA expedition transect, thereby broadening the knowledge of UCYN-A presence in significant numbers to previously under-sampled regions such as the Indian Ocean (Fig. 1). These findings underscore the fact that the known distribution of this unusual and versatile diazotroph is continuing to expand, which calls for a continuous re-evaluation of the significance of UCYN-A to global marine N_2 fixation.

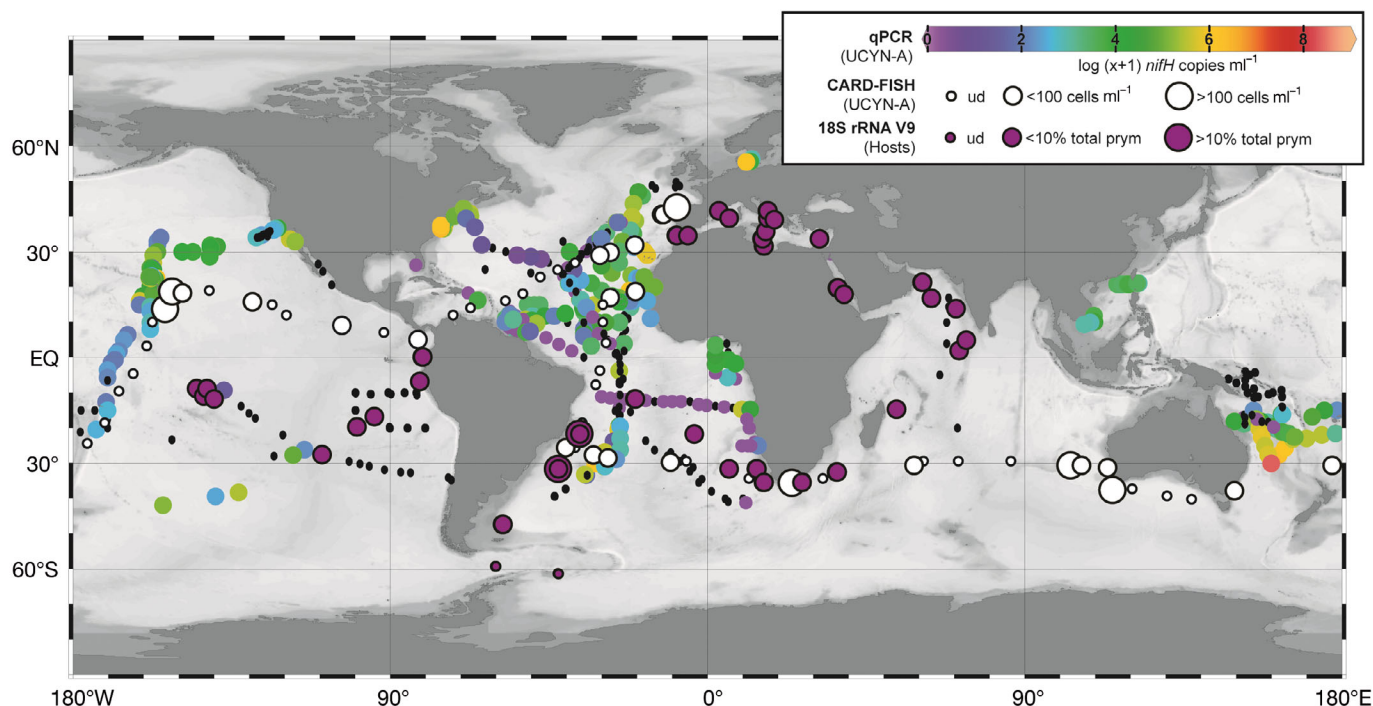


Fig. 1. Global distribution of *Atelocyanobacterium thalassa* (UCYN-A). Compilation of all published UCYN-A association abundance data acquired using TaqMan[®] qPCR, catalyzed reporter deposition-fluorescence *in situ* hybridization (CARD-FISH), and variable region tag 18S rRNA gene-based approaches. Table S2 in Supplement 2 at www.int-res.com/articles/suppl/a077p125_supp.xlsx summarizes the qPCR-based studies included. CARD-FISH and 18S rRNA gene V9 tag data were adapted from Cabello et al. (2016). For each station, the maximum UCYN-A abundances reported using any of the TaqMan[®] based assays, converted to $\log(x+1)$ *nifH* copies l^{-1} , are displayed as colored dots. Small purple dots: UCYN-A below quantification limits of the qPCR assay used in the source study; small black dots: UCYN-A was undetected. CARD-FISH and 18S rRNA gene V9 tag data from Cabello et al. (2016) are overlaid using white and purple circles, respectively

RESPONSE OF UCYN-A TO ENVIRONMENTAL VARIABLES AND NUTRIENT ADDITION MANIPULATIONS

The distribution and activity of diazotrophs is believed to be controlled by a number of parameters including temperature, nitrogen-to-phosphorus (N:P) ratios, and iron (Fe) (Goebel et al. 2010, Moisander et al. 2010, Ward et al. 2013). Currently, the nutrients limiting UCYN-A growth and the other factors that control abundances such as grazing and viral lysis are not well understood (Sohm et al. 2011). In a study in the South Pacific Ocean, UCYN-A abundances increased when either organic C or Fe were added to incubation experiments, and maximum net growth rates of $0.19\ d^{-1}$ were reported (Moisander et al. 2012). In contrast, P alone did not have an impact on UCYN-A growth (Moisander et al. 2012). In the North Atlantic Ocean, aerosol input of Fe may relieve diazotrophs from Fe limitation, and indeed UCYN-A abundances were recently shown to correlate with input of North African aerosols (Ratten et al. 2015). In a previous study in the North Atlantic Ocean, UCYN-

A *nifH* expression increased in P addition treatments (Turk-Kubo et al. 2012). However, despite increases in UCYN-A *nifH* expression, P additions did not enhance N_2 fixation rates or UCYN-A abundances (Turk-Kubo et al. 2012). The presence of biologically available N is thought to inhibit nitrogenase activity (Dixon & Kahn 2004). However, substantial N_2 fixation in N-replete waters (Knapp 2012) and lack of inhibition of N_2 fixation activity in nitrate amended cultures of the unicellular diazotrophic cyanobacterium *Crocospaera* (Dekazemacker & Bonnet 2011) indicate that the regulation of N_2 fixation activity and growth of diazotrophs may be more complex. In a bioassay experiment in the North Atlantic, UCYN-A *nifH* abundances increased in both ammonium nitrate and ammonium nitrate and phosphorus treatments (Langlois et al. 2012). Interestingly, near the Cape Verde islands, inhibition of UCYN-A *nifH* expression and N_2 fixation was not observed when incubations were amended with ammonium nitrate (Krupke et al. 2015). In addition, although neither UCYN-A *nifH* expression nor bulk N_2 fixation measurements showed a clear response to nutrient amend-

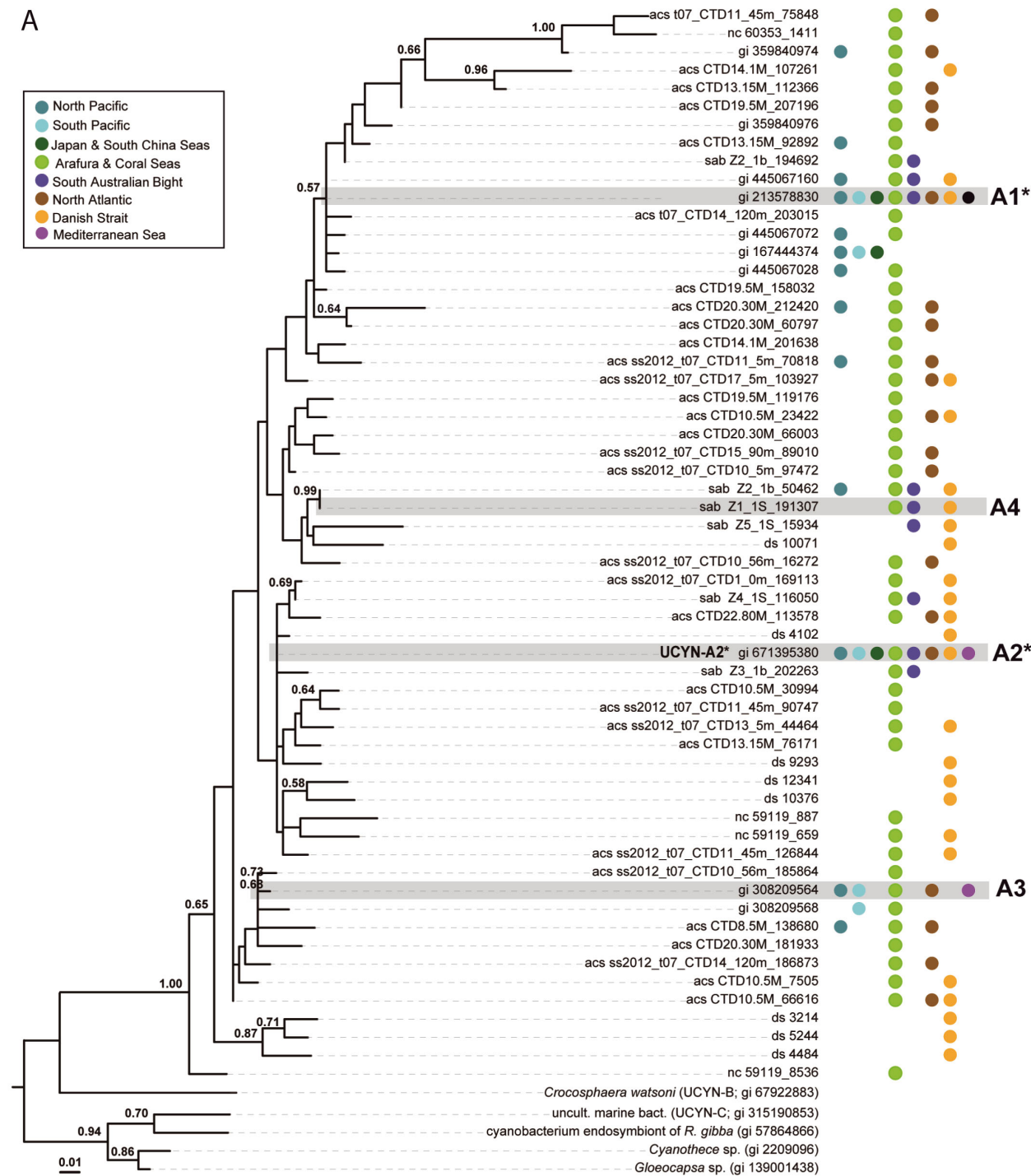
ments, nanoSIMS showed that single-cell UCYN-A N_2 and CO_2 fixation differed between treatments with a positive response of N_2 fixation activity to Fe and P additions (Krupke et al. 2015). Taken together, these results illustrate the complexity of nutrient limitations in diazotrophs and the difficulty of interpreting links between species-specific *nifH* gene expression and N_2 fixation rates from a mixed community.

INCREASING AMOUNTS OF UCYN-A *nifH* SEQUENCES REVEAL NOVEL SUBLINEAGES

Genetic diversity has been linked to identification of niche-specific ecotypes within several highly abundant marine bacteria such as *Synechococcus* and *Prochlorococcus*, as well as the less abundant unicellular N_2 -fixing cyanobacterium *Crocosphaera* (e.g. Webb et al. 2009, Coleman & Chisholm 2010, Bench et al. 2013, Kashtan et al. 2014, Sohm et al. 2016). UCYN-A was once thought to be an organism of broad distribution but potentially low genetic diversity (Tripp et al. 2010). However, because of the standard of analyzing amino acid sequences as an indicator of diversity for functional genes, variation on the nucleic acid level of the UCYN-A *nifH* gene sequences was for a long time undetected. Although almost identical at the amino acid level, UCYN-A *nifH* nucleic acid sequences cluster into several sublineages, and based on this observation, Thompson et al. (2014) defined 3 distinct sublineages: UCYN-A1, UCYN-A2 and UCYN-A3. The UCYN-A1 sublineage is represented by the *nifH* sequence of the first UCYN-A genome sequenced, sampled from Stn ALOHA in the North Pacific Subtropical Gyre (Zehr et al. 2008, Tripp et al. 2010). The UCYN-A2 sublineage is represented by the *nifH* sequence of a genome sequenced from a population of UCYN-A cells sampled from the Scripps Institute of Oceanography (SIO) pier at La Jolla, California, USA (Bombar et al. 2014). Thompson et al. (2014) also defined a third sublineage, UCYN-A3, but a representative of this sublineage has not yet had its genome sequenced. Genome sequencing of UCYN-A1 and UCYN-A2 revealed that these 2 sublineages were highly similar in gene content, suggesting that they are adapted to a similar lifestyle (Bombar et al. 2014). However, despite the high similarity in gene content between the 2 genomes, the shared genes were on average only 86% similar on the amino acid level, indicating that species divergence has occurred from a common ancestor (Bombar et al. 2014, Cornejo-Castillo et al. 2016). Recent CARD-FISH observations suggest high fidelity and co-evolution

between UCYN-A1 and UCYN-A2 and the respective eukaryotic partner for each sublineage (Cabello et al. 2016, Cornejo-Castillo et al. 2016). However, as these techniques have been used in relatively few studies, the question of host fidelity remains open. As with other marine cyanobacteria, the co-occurrence of multiple UCYN-A sublineages and/or forming an association with multiple eukaryotic partner cells may have significant impacts in the response to different environmental conditions, and consequently N_2 fixation activity.

In recent years, *nifH* amplicon sequencing using next generation sequencing technologies have greatly increased the number of *nifH* sequences (e.g. Farnelid et al. 2011, Bentzon-Tilia et al. 2015, Messer et al. 2015, 2016, Turk-Kubo et al. 2015, Xiao et al. 2015, Cheung et al. 2016). Compared to traditional clone library studies, sublineages present at low abundances are more likely to be recovered using next generation sequencing, which aids our efforts to describe novel UCYN-A sublineages. To investigate this, we compiled *nifH* UCYN-A sequences from GenBank (available as of 17 December 2015) as well as those reported from next generation sequencing studies (Fig. 2). Briefly, raw read data from the Danish Strait (Bentzon-Tilia et al. 2015), South Australian Bight (Messer et al. 2015) and Arafura and Coral Seas (Messer et al. 2016), including the New Caledonia lagoon (Turk-Kubo et al. 2015), were obtained from the Sequence Read Archive (SRA) at the National Center for Biotechnology Information (NCBI). Raw reads were merged, when possible, using PEAR (Zhang et al. 2014), quality filtered using a phred score of 20 in QIIME (Caporaso et al. 2010), and chimeras were removed using the 'usearch61' de novo chimera check algorithm (Edgar et al. 2011). Potential UCYN-A sequences were identified via clustering with UCYN-A sequences from Thompson et al. (2014) using CD-HIT-EST or CD-HIT-EST-2D (Huang et al. 2010) at 90% nt similarity. Sequences with stop codons or gaps were removed from the analysis after import into ARB (Ludwig et al. 2004) and translation into amino acid sequences. A majority of reads from the South Australian Bight were removed due to stop codons and/or gaps. Neighbor joining trees were constructed in ARB to identify the few representative sequences that had >90% nt similarity to UCYN-A but were more closely related to other unicellular cyanobacteria such as *Cyanothece*. Data from Farnelid et al. (2011) were not included due to the short sequence length resulting from early 454-pyrosequencing studies, and data from Xiao et al. (2015) were not included because <25



B

Rep. Seq. name	total number of reads in ngs datasets					total number of Genbank sequences from each region				
	New Caledonia (nc)	Arafura & Coral Seas (acs)	South Australian Bight (sab)	Danish Strait (ds)	North Pacific	South Pacific	Japan & South China Seas	North Atlantic	Danish Strait	MedSea
gi 671395380 (UCYN-A2)	215549	4546	282	18429	86	5	7	53	25	13
gi 213578830 (UCYN-A1)	41187	9146	288	118	273	13	12	18		
sab Z1_1S_191307 (UCYN-A4)	14	1	117	5763						
gi 308209564 (UCYN-A3)	481	879			2	6		25		3

total UCYN-A reads were identified. When clustered at 97 % nucleic acid similarity (CD-HIT-EST), and after removing clusters that comprised less than 0.1 % of the total UCYN-A sequences from a given study, 59 *nifH* UCYN-A operational taxonomic units (OTUs) were identified (Fig. 2, Table S2 in Supplement 2 at www.int-res.com/articles/suppl/a077p125_supp.xlsx). As expected, the next generation sequencing datasets greatly increased the number of *nifH* UCYN-A sequences over those present in the nr database of GenBank (GenBank: 819 reads; Turk-Kubo et al. 2015: 268943 reads; Messer et al. 2015: 781 reads; Messer et al. 2016: 18331 reads; Bentzon-Tilia et al. 2015: 25841 reads; Table S2 in Supplement 2). The large numbers of sequences produced by next generation platforms provide an emerging pattern of UCYN-A *nifH* sublineages. Notably, due to the high similarity between the UCYN-A sequences, the previously identified sublineages are not well supported by bootstrap values in this analysis (Fig. 2A). To allow for comparisons in geographic distribution, relative abundances and qPCR assay specificity we therefore chose to focus on the dominating OTU within each sublineage, referred to here as the representative OTU (Fig. 2B).

In agreement with GenBank sequences, a majority of the next generation sequences affiliated with UCYN-A1 and UCYN-A2, and are reported from diverse ocean basins worldwide (Fig. 2A, Table S2). UCYN-A2 sequences frequently co-occur with UCYN-A1. However in some studies, UCYN-A2 sequences have been detected while UCYN-A1 sequences were not; e.g. the SIO Pier in the Gulf of Santa Catalina (Thompson et al. 2014), Heron Reef in the Coral Sea (Hewson et al. 2007) and the eastern Mediterranean Sea (Man-Aharonovich et al. 2007). Out of all UCYN-A *nifH* sequences in GenBank, the representative sequences of UCYN-A1 and UCYN-A2 correspond to 58.4 and 23.1 % of the sequences, respectively (Table S2). Interestingly, the relative abundance of the UCYN-A2 representative OTU

dominated next generation sequencing studies from the Coral Sea (80.1 %; Turk-Kubo et al. 2015) and the Danish Strait (71.3 %; Bentzon-Tilia et al. 2015). In addition, the UCYN-A2 representative OTU was consistently present in the next generation sequencing studies (Fig. 2). This illustrates that the total number of sequences recovered from each UCYN-A sublineage is likely to be largely biased by the sampling location and sequencing depth used in the study. Sequences affiliated with the UCYN-A3 sublineage have previously been reported from near the Cape Verde Islands in the tropical North Eastern Atlantic (G. Wheeler, direct submission to GenBank) at Stn ALOHA in the North Pacific (Zehr et al. 2007), in the eastern Mediterranean Sea (Man-Aharonovich et al. 2007, Le Moal et al. 2011) and the South Pacific Gyre (Halm et al. 2012). The UCYN-A3 representative OTU corresponds to 4.4 % of the total *nifH* UCYN-A sequences in GenBank (Fig. 2B, Table S2). Within the next generation sequencing datasets, the representative UCYN-A3 sequence type was present at relative abundances of 4.8 and 0.2 % in the Arafura and Coral Seas (Messer et al. 2016) and the Noumea Lagoon in New Caledonia (Turk-Kubo et al. 2015, Table S2), respectively. Thus, the UCYN-A3 sublineage also appears to have a wide global distribution and can co-occur with other UCYN-A sublineages.

Several UCYN-A *nifH* OTUs form a cluster with a small number of previously reported sequences in GenBank from the North Pacific (accession numbers LC013598, LC013602, LC013603 and LC013607; Shiozaki et al. 2015), the northeast Atlantic (accession numbers KF546393, KF546401, KF546415, KF546453 and KF546389; G. Wheeler unpubl. data) and Cape Verde islands in the eastern Atlantic Ocean (accession number HQ634491; S. Sudhaus & J. LaRoche unpubl. data). Interestingly, sequences within this sublineage were present in all 4 of the next generation sequencing datasets investigated here and corresponded to a relative abundance of 22.3 % in the Danish Strait (Fig. 2, Table S2). Thus,

Fig. 2. (A) Maximum likelihood (ML) tree of *Atelocyanobacterium thalassa* (UCYN-A) partial *nifH* nucleotide sequences (313 bp), constructed using operational taxonomic units (OTUs) generated from all UCYN-A sequences as available in GenBank (17 December 2015) and next generation sequencing based studies known to date. Phylogenetic analysis was conducted in MEGA 6.0. Branch lengths were determined using the Tamura-Nei model, bootstrap values were calculated using 1000 replicate trees and are displayed when greater than 50. The ML tree and associated metadata were visualized using the iTOL v2 web platform (Letunic & Bork 2011). The presence (colored circle) or absence (no circle) of UCYN-A OTUs in each oceanic region is described to the right of the tree. OTUs that represent dominant sequence types for each sublineage are indicated with grey shading and labeled A1, A2, A3 and A4. The OTUs containing the genome sequenced UCYN-A *nifH* are indicated with an asterisk. (B) Total number of UCYN-A *nifH* reads available from GenBank and next generation sequencing (ngs) datasets for different oceanic regions for each of the sublineage OTU representatives (indicated with grey shading in A). The representative sequences are indicated with origin New Caledonia Lagoon (nc; Turk-Kubo et al. 2015), Arafura & Coral Seas (acs; Messer et al. 2016), South Australian Bight (sab; Messer et al. 2015), and the Danish Strait (ds; Bentzon-Tilia et al. 2015)

despite the presence of only a few representatives of this sublineage in the GenBank nr database, next generation sequencing-based studies indicate that this novel UCYN-A sublineage, which we refer to henceforth as UCYN-A4, may be significant in some environments. Little is known about its global distribution at this time, and more studies that employ next generation sequencing will be required to better characterize their distribution. However, the presence of this sublineage in vastly different oceanic regions suggests that UCYN-A4 may have a wide geographic distribution, similar to other UCYN-A sublineages. Furthermore, based on the next generation sequencing data from the Danish Strait (Bentzon-Tilia et al. 2015), it appears likely that UCYN-A4 may be present at relatively high abundances in specific areas, yet since UCYN-A4 *nifH* sequences have largely evaded detection within clone library efforts, they may often be present at lower abundances compared to other UCYN-A sublineages.

One of the challenges emerging from the discovery of multiple UCYN-A sublineages is evaluating the specificity and cross-hybridization of molecular assays. Recently, the most widely used qPCR assay used to quantify UCYN-A, designed by Church et al. (2005a) (Table S1), was found to contain several mismatches to *nifH* sequences representing UCYN-A2 and UCYN-A3 sublineages (Thompson et al. 2014, Table 1). Similarly, in a compilation of all UCYN-A qPCR assays reported, we found that each of the primer/probe sets contained several mismatches to one or more of the UCYN-A sublineage OTU representatives reported in this study (Table 1). Thus, since UCYN-A sublineages frequently co-occur, *nifH* abundance and transcription may have been under-

estimated in numerous studies if qPCR using only 1 assay was the only method for investigating UCYN-A presence and activity. To circumvent this problem, a qPCR assay targeting UCYN-A2 was designed by Thompson et al. (2014), which has been used in combination with the Church et al. (2005a) qPCR assay in several subsequent studies with the aim to evaluate abundances of the respective UCYN-A sublineages and to estimate the total abundance of UCYN-A (Bombar et al. 2014, Bonnet et al. 2015, Turk-Kubo et al. 2015, Rees et al. 2016). Interestingly, using the qPCR assay designed to target UCYN-A2 (5 total mismatches to UCYN-A1; Table 1), high abundances (up to 2×10^5 *nifH* gene copies l⁻¹) were reported from the SIO pier in La Jolla, California (Thompson et al. 2014). Similarly, Bentzon-Tilia et al. (2015) measured UCYN-A abundances of up to 7.6×10^5 *nifH* gene copies l⁻¹ using a qPCR assay with a total of 5 mismatches to UCYN-A1 in 2 temperate estuaries in the Danish Strait. These results suggest that sublineages that would not have been detected using the Church et al. (2005a) qPCR assay may be prevalent in some environments and could account for significant N₂ fixation. Unfortunately, due to the short *nifH* fragment and the high similarity between the different sublineages (Fig. 2A), designing a qPCR assay to differentiate between UCYN-A2 and UCYN-A3 is difficult, and *in silico* tests indicate that the UCYN-A2 qPCR assay designed by Thompson et al. (2014) does not contain sufficient mismatches to prevent cross-hybridization with the UCYN-A3 sublineage (Table 1). Consequently, with the increasing number of UCYN-A *nifH* sequences and the discovery of closely related sublineages, it will be important to re-evaluate the amplification specificity and cover-

Table 1. Number of total mismatches of primers and probes of all published qPCR assays designed to quantify *Atelocyanobacterium thalassa* (UCYN-A) to the operational taxonomic unit (OTU) representative of each UCYN-A sublineage

qPCR assay name (reference sequence if available)	Number of total mismatches of primers and probes to the OTU representative of each sublineage				Reference
	UCYN-A1 (gi 213578830)	UCYN-A2 (gi 671395380)	UCYN-A3 (gi 308209564)	UCYN-A4 (sab Z1_1S_191307)	
OTU0002	5	0	0	0	Bentzon-Tilia et al. (2015)
Group A (AF059642)	2	5	4	4	Church et al. (2005a)
UCYN-A (HM210392)	1	1	0	2	Halm et al. (2012)
UCYN-A- <i>nifH</i> -2	0	0	2	1	Hashimoto et al. (2016)
UCYN-A (HQ456051)	0	0	2	2	Kong et al. (2011) ^a
Group A (AF059627)	0	4	5	4	Langlois et al. (2008)
Group A (AB665496)	0	6	7	5	Shiozaki et al. (2014)
CBga (CB4713A1)	6	0	1	1	Short & Zehr (2007)
UCYN-A2	5	0	1	1	Thompson et al. (2014)

^aThe qPCR assay by Kong et al. (2011) is SYBR-based and uses the reverse primer by Church et al. (2005a)

age of UCYN-A *nifH* qPCR assays before making interpretations of ecotype distributions and abundances. In addition, evaluating the UCYN-A *nifH* sublineages present in a sample using an amplicon library is necessary in order to evaluate which qPCR assays should be used for estimating total UCYN-A abundances within a sample. The wide distribution and co-occurrence of UCYN-A sublineages suggests that if the UCYN-A sublineages represent distinct ecotypes they have overlapping niches. Thus, more information, including highly specific assays, possibly beyond the *nifH* gene fragment currently used, will be needed in order to evaluate what controls their spatial and temporal distribution patterns.

VISUALIZATION TECHNIQUES ALLOW FOR INSIGHTS INTO CELL MORPHOLOGY AND PHYSIOLOGY

Recent developments of UCYN-A-specific FISH probes have led to the first visualizations of UCYN-A cells and have resulted in a rapid increase in knowledge of the morphology and physiology of this unusual symbiosis. Early CARD-FISH studies investigating unicellular diazotrophs used the Nitro821R probe (Mazard et al. 2004), which targets the 16S rRNA gene of unicellular cyanobacterial diazotrophs. Using this approach, physiologically diverse species including *Crocospaera* sp., *Cyanothece* sp., intracellular cyanobacterial symbionts of the marine diatom *Climacodium frauenfeldianum* as well as UCYN-A would be hybridized at the same time. In concurrence with *nifH* amplicon-based studies, cells that hybridized with the Nitro821R probe were numerically significant at times, indicating the importance of this group of diazotrophs in diverse marine locations including the Pacific Ocean (Bonnet et al. 2009), the Atlantic Ocean (Agawin et al. 2014, Riou et al. 2016), the New Caledonia coral lagoon (Biegala & Raimbault 2008) and the Mediterranean Sea (Le Moal & Biegala 2009, Le Moal et al. 2011). However, although specific morphological features and presence within various size fractions were observed, the lack of specificity of the probe and the co-occurrence of several unicellular cyanobacterial diazotrophs precluded links to the various species.

The large genome reduction (Tripp et al. 2010, Bombar et al. 2014) and tight coupling in nutrient exchange suggests an obligate symbiosis between UCYN-A and its associated eukaryotic partner (Krupke et al. 2015). Yet, the association is fragile and without careful sample processing, UCYN-A

may be easily dislodged from the eukaryotic cell (Thompson et al. 2012, Krupke et al. 2013). Epifluorescence and scanning electron microscopy observations suggest that the UCYN-A cell may be connected to the surface of the eukaryotic partner in a loosely attached, yet intimate cell-surface interaction between the symbionts (Thompson et al. 2012, Krupke et al. 2013, 2014, 2015, Cabello et al. 2016, Cornejo-Castillo et al. 2016). However, using transmission electron microscopy (TEM), spherical bodies within *Braarudosphaera bigelowii* cells, separated from the eukaryotic cell cytoplasm by a single membrane, were identified as UCYN-A endosymbionts, suggesting that the association could also be intracellular (Hagino et al. 2013).

There are now several UCYN-A specific FISH probes that have been developed and applied to environmental samples, which has advanced the ability to link morphotypes with 16S rRNA gene phylogenies. Thus far, the most frequently reported observation with UCYN-A-specific probes (UCYN-A732 and/or UCYN-A159) is one UCYN-A cell per eukaryotic partner cell located on the polar end of the eukaryotic cell (Thompson et al. 2012, Krupke et al. 2013, 2015, Cabello et al. 2016). Interestingly, in a small fraction of observed associations, 2 hybridized UCYN-A cells have been reported on opposite ends of a single eukaryotic cell (Thompson et al. 2012, Cabello et al. 2016).

After the description of a second UCYN-A sublineage, UCYN-A2 (Thompson et al. 2014), Cornejo-Castillo et al. (2016) designed a double CARD-FISH assay to specifically target the 2 different UCYN-A sublineages. In the South Atlantic, they observed small prymnesiophyte cells associated with a single cell hybridized with the UCYN-A1 probe and multiple UCYN-A2 (~3 to 10 cells) within a cellular pocket of the eukaryotic partner cell. Similar morphological observations of UCYN-A1 and UCYN-A2 have been observed at Stn ALOHA in the subtropical Pacific Ocean and at the SIO pier (Fig. 3). In comparing the genomes between UCYN-A1 and UCYN-A2, several genes determining cell shape and cell wall biogenesis were present in UCYN-A2 while absent in UCYN-A1, suggesting differences in how the 2 sublineages are associated with their eukaryotic partner cells (Bombar et al. 2014). Thus, for some UCYN-A sublineages, multiple cells may be associated with one eukaryotic partner cell, but how cell division occurs within these cell consortia is still not clear (Zehr 2015).

The eukaryotic UCYN-A partner cells have been visualized using several different CARD-FISH assays (Krupke et al. 2014, Cabello et al. 2016, Cornejo-

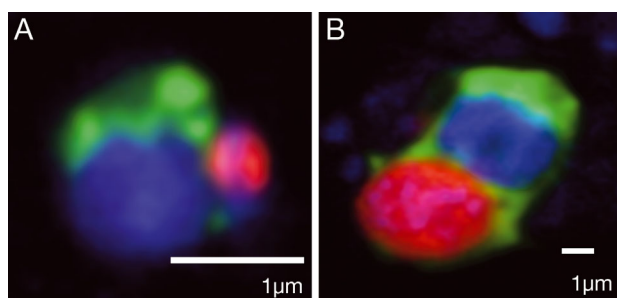


Fig. 3. Epifluorescence micrographs of *Atelocyanobacterium thalassa* (UCYN-A) (red) and its symbiotic partner (green) stained with catalyzed reporter deposition-fluorescence *in situ* hybridization (CARD-FISH) and DAPI at the Scripps Institute of Oceanography (SIO) pier, La Jolla, CA, USA. The nucleus of each host is visualized with DAPI stain (blue). (A) UCYN-A1 and its prymnesiophyte partner using UCYN-A1 732 and UPRYM-69 probes (Cornejo-Castillo et al. 2016). (B) UCYN-A2 targeted by the UCYN-A2 732 probe and the larger prymnesiophyte with the UBRADO69 probe (Cornejo-Castillo et al. 2016). Micrographs were obtained using Leica SP5 Confocal Microscope at the University of California, Santa Cruz Life Sciences Microscopy Center

Castillo et al. 2016). CARD-FISH, in combination with 18S rRNA tag sequencing, revealed UCYN-A in association with a small prymnesiophyte cell in all ocean basins, and suggests that the eukaryotic partner cells of UCYN-A1 and A2 are also widely distributed (Cabello et al. 2016, our Fig. 1). Based on observations of different sizes of the hosts, as well as recruitment of genomes from metagenomes of different size fractions, the cell consortia of UCYN-A1 appears to be smaller in size compared to UCYN-A2 (Thompson et al. 2014, Cabello et al. 2016, Cornejo-Castillo et al. 2016). Previously, Thompson et al. (2014) estimated the size of the UCYN-A2 cell consortia using microscopy and flow cytometry to be ~7 to 10 µm in diameter. Interestingly, Hagino et al. (2013) observed *B. bigelowii* cells, that had an 18S

rRNA sequence identical to that of the UCYN-A2-associated eukaryotic cells described by Thompson et al. (2014), to be 15 µm in diameter with pentaliths. Together, these observations suggest that the eukaryotic partner cells may have different morphological stages and the consortia could thereby vary in size.

In addition to the differences in size of the eukaryotic partner cells, differences in the size of UCYN-A cells have also been observed. In the North Pacific and North Atlantic, small UCYN-A cells (0.5 to 1.2 µm in diameter) associated with small eukaryotic cells (1 to 3 µm in diameter) were found (Thompson et al. 2012, Krupke et al. 2014, 2015, our Table 2). In the North Atlantic, slightly larger associations have also been described, with UCYN-A cells of up to 3 µm in diameter with a eukaryotic cell of up to 5.5 µm in diameter co-occurring with the smaller, previously described UCYN-A consortia (Krupke et al. 2013). Interestingly, although these differences in cell size could represent cells in different stages of cell division, the samples that contained multiple size ranges of UCYN-A cells were collected near the Cape Verde islands, a region where sequences from multiple UCYN-A sublineages have been reported and where UCYN-A is thought to be responsible for the majority of the biological N₂ fixation (Goebel et al. 2010, Turk et al. 2011). Unfortunately, it is not yet resolved how CARD-FISH observations can be linked to the defined UCYN-A *nifH* sublineages, since the 16S rRNA gene sequences of UCYN-A3 and UCYN-A4 sublineages are unknown (current knowledge of the different sublineages is summarized in Table 3). However, the known variation in cell sizes of the UCYN-A1 and UCYN-A2 sublineages and their eukaryotic partner cells suggests differences in physiology and nutrient requirements. In a recent metatranscriptome study, the number of *nifH* transcripts per UCYN-A2 cell were almost double compared to UCYN-A1, despite UCYN-A1 being

Table 2. The diameter of *Atelocyanobacterium thalassa* (UCYN-A) and associated eukaryotic cells estimated using halogenated *in situ* hybridization (HISH) or catalyzed reporter deposition-fluorescence *in situ* hybridization (CARD-FISH) assays and the respective probes used in each study. na: not available. The study by Cornejo-Castillo et al. (2016) has not been included because the measured diameters of the cells were not reported

UCYN-A cell diameter (µm)	N	Eukaryotic cell diameter (µm)	N	UCYN-A probe	Eukaryotic probe	Reference
0.62 ± 0.16	19	1.36 ± 0.24	10	UCYN-A732	Not used	Thompson et al. (2012)
1.36 ± 0.45	44	3.58 ± 1.42	8	UCYN-A732, UCYN-A159	Not used	Krupke et al. (2013)
1.16 ± 0.24	52	2.26 ± 0.75	52	UCYN-A732	PRYM02	Krupke et al. (2014)
0.83 ± 0.15	44	1.66 ± 0.23	44	UCYN-A732	PRYM02	Krupke et al. (2015)
na		~2.5	2030	UCYN-A732	UPRYM-69 (UCYN-A1 partner)	Cabello et al. (2016)
na		4-5	17	UCYN-A732	UBRADO-69 (UCYN-A2 partner)	Cabello et al. (2016)

Table 3. Synthesis of available sequence information, assays available to detect or visualize *Atelocyanobacterium thalassa* (UCYN-A), and known morphology of UCYN-A sublineages. CARD-FISH: catalyzed reporter deposition-fluorescence *in situ* hybridization

	UCYN-A1	UCYN-A2	UCYN-A3	UCYN-A4
Sequence information				
Full genome	Yes	Yes	No	No
<i>nifH</i> gene	Yes	Yes	Yes	Yes
16S rRNA gene	Yes	Yes	Unknown	Unknown
18S rRNA gene from eukaryotic partner	Yes	Yes	Unknown	Unknown
Assays				
<i>nifH</i> qPCR assay	Yes, not specific	Yes, not specific	Yes, not specific	Yes, not specific
16S rRNA gene probe (CARD-FISH)	Yes	Yes	Unknown	Unknown
18S rRNA gene probe (CARD-FISH)	Yes	Yes	Unknown	Unknown
Known morphology				
Number of cells per eukaryotic partner	1–2	~3–10	Unknown	Unknown
Size of consortia (μm)	1–3	4–10	Unknown	Unknown

more abundant in the sample, possibly reflecting the additional nutrient requirements of the larger eukaryotic partner (Cornejo-Castillo et al. 2016). Identification and visualization of additional UCYN-A sublineages will therefore be important for understanding their physiology and to estimate their contribution to marine N_2 fixation.

FUTURE DIRECTIONS AND CONCLUSIONS

Because of the low abundances of UCYN-A compared to other marine microorganisms, UCYN-A *nifH* sequences were rarely detected within early metagenomic datasets (Johnston et al. 2005). With the availability of the UCYN-A1 and UCYN-A2 genomes and the 18S rRNA gene sequences of the corresponding eukaryotic partners, in combination with a great increase in sequencing depth, screening of metagenomic and metatranscriptomic datasets for UCYN-A have become more successful. Within the metagenomic datasets from the Tara-Oceans Expedition, UCYN-A-associated prymnesiophytes appeared to be consistently present (albeit at low relative abundances) in widely distributed regions, including areas where UCYN-A diversity and abundance are not yet well known (Cabello et al. 2016, our Fig. 1). In addition, large fractions of UCYN-A1 and UCYN-A2 genomes (genome recovery of close to 100% for some samples) were recently recruited from metagenomes and metatranscriptomes from 2 stations in the South Atlantic Ocean (Cornejo-Castillo et al. 2016). Thus, with the increasing amounts of sequencing data available and target genes beyond the *nifH* gene, knowledge of the global distribution patterns,

transcriptional activity and abundance of this symbiosis is expected to rapidly increase in the near future.

Over the past decade, multiple novel methods and approaches have been used to study UCYN-A, which has led to remarkable discoveries about this unusual symbiosis. In the future, it will be important to use cross comparisons for validation of measurements using different assays and methods. For example, current qPCR methods result in an order of magnitude higher abundances of UCYN-A compared to CARD-FISH counts (Krupke et al. 2013), and the evaluation of primer specificity for different UCYN-A sublineages of both qPCR and CARD-FISH assays and how they compare have not yet been resolved. Based on the evidence of diverse lifestyles of different UCYN-A sublineages, designing novel assays to distinguish between the sublineages will be important for elucidating UCYN-A ecotype diversity and distribution. Furthermore, renewed efforts to isolate this association are needed to make breakthroughs in understanding the biology of this unusual symbiosis, as there remain many unanswered questions about the structural basis of the symbiosis, cell division and the communication between partner cells.

Acknowledgements. We thank Lauren Messer for providing metadata from her sample sites, and Benjamin Abrams and the UCSC Life Science Microscopy Center for microscopy assistance. H.F. was supported by the Swedish Research Council VR 637-2013-7502, and H.F., K.T.-K. and J.P.Z. were supported through the NSF C-MORE (#EF0424599) and the Simons Collaboration on Ocean Processes and Ecology (SCOPE). M.C.M.-M. was supported by a Marie Curie International Outgoing Fellowship within the 7th European Community Framework Programme.

LITERATURE CITED

- Agawin NSR, Benavides M, Busquets A, Ferriol P, Stal LJ, Aristegui J (2014) Dominance of unicellular cyanobacteria in the diazotrophic community in the Atlantic Ocean. *Limnol Oceanogr* 59:623–637
- Benavides M, Agawin NSR, Aristegui J, Ferriol P, Stal LJ (2011) Nitrogen fixation by *Trichodesmium* and small diazotrophs in the subtropical northeast Atlantic. *Aquat Microb Ecol* 65:43–53
- Benavides M, Moisander PH, Daley MC, Bode A, Aristegui J (2016) Longitudinal variability of diazotroph abundances in the subtropical North Atlantic Ocean. *J Plankton Res* 38:662–672
- Bench SR, Heller P, Frank I, Arciniega M, Shilova IN, Zehr JP (2013) Whole genome comparison of six *Crocospaera watsonii* strains with differing phenotypes. *J Phycol* 49: 786–801
- Bentzon-Tilia M, Traving SJ, Mantikci M, Knudsen-Leerbeck H, Hansen JL, Markager S, Riemann L (2015) Significant N₂ fixation by heterotrophs, photoheterotrophs and heterocystous cyanobacteria in two temperate estuaries. *ISME J* 9:273–285
- Biegala IC, Raimbault P (2008) High abundance of diazotrophic picocyanobacteria (<3 µm) in a Southwest Pacific coral lagoon. *Aquat Microb Ecol* 51:45–53
- Bombar D, Heller P, Sanchez-Baracaldo P, Carter BJ, Zehr JP (2014) Comparative genomics reveals surprising divergence of two closely related strains of uncultivated UCYN-A cyanobacteria. *ISME J* 8:2530–2542
- Bonnet S, Biegala IC, Dutrieux P, Slemons LO, Capone DG (2009) Nitrogen fixation in the western equatorial Pacific: rates, diazotrophic cyanobacterial size class distribution, and biogeochemical significance. *Global Biogeochem Cycles* 23:GB3012
- Bonnet S, Rodier M, Turk-Kubo KA, Germineaud C and others (2015) Contrasted geographical distribution of N₂ fixation rates and *nifH* phylotypes in the Coral and Solomon Seas (southwestern Pacific) during austral winter conditions. *Global Biogeochem Cycles* 29: 1874–1892
- Brown SM, Jenkins BD (2014) Profiling gene expression to distinguish the likely active diazotrophs from a sea of genetic potential in marine sediments. *Environ Microbiol* 16:3128–3142
- Cabello AM, Cornejo-Castillo FM, Raho N, Blasco D and others (2016) Global distribution and vertical patterns of a prymnesiophyte–cyanobacteria obligate symbiosis. *ISME J* 10:693–706
- Capone DG, Zehr JP, Paerl HW, Bergman B, Carpenter EJ (1997) *Trichodesmium*, a globally significant marine cyanobacterium. *Science* 276:1221–1229
- Capone DG, Burns JA, Montoya JP, Subramaniam A and others (2005) Nitrogen fixation by *Trichodesmium* spp.: an important source of new nitrogen to the tropical and subtropical North Atlantic Ocean. *Global Biogeochem Cycles* 19:GB2024
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K and others (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336
- Cheung S, Xia X, Guo C, Liu H (2016) Diazotroph community structure in the deep oxygen minimum zone of the Costa Rica Dome. *J Plankton Res* 38:380–391
- Church MJ, Jenkins BD, Karl DM, Zehr JP (2005a) Vertical distributions of nitrogen-fixing phylotypes at Stn ALOHA in the oligotrophic North Pacific Ocean. *Aquat Microb Ecol* 38:3–14
- Church MJ, Short CM, Jenkins BD, Karl DM, Zehr JP (2005b) Temporal patterns of nitrogenase gene (*nifH*) expression in the oligotrophic North Pacific Ocean. *Appl Environ Microbiol* 71:5362–5370
- Church MJ, Mahaffey C, Letelier RM, Lukas R, Zehr JP, Karl DM (2009) Physical forcing of nitrogen fixation and diazotroph community structure in the North Pacific subtropical gyre. *Global Biogeochem Cycles* 23:GB2020
- Coleman ML, Chisholm SW (2010) Ecosystem-specific selection pressures revealed through comparative population genomics. *Proc Natl Acad Sci USA* 107:18634–18639
- Cornejo-Castillo FM, Cabello AM, Salazar G, Sa P and others (2016) Cyanobacterial symbionts diverged in the late Cretaceous towards lineage-specific nitrogen fixation factories in single-celled phytoplankton. *Nat Commun* 7: 11071
- Dekazemacker J, Bonnet S (2011) Sensitivity of N₂ fixation to combined nitrogen forms (NO₃ and NH₄) in two strains of the marine diazotroph *Crocospaera watsonii* (Cyanobacteria). *Mar Ecol Prog Ser* 438:33–46
- Dixon R, Kahn D (2004) Genetic regulation of biological nitrogen fixation. *Nat Rev Microbiol* 2:621–631
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27:2194–2200
- Falcón LI, Carpenter EJ, Cipriano F, Bergman B, Capone DG (2004) N₂ Fixation by unicellular bacterioplankton from the Atlantic and Pacific Oceans: phylogeny and in situ rates. *Appl Environ Microbiol* 70:765–770
- Farnelid H, Andersson AF, Bertilsson S, Al-Soud WA and others (2011) Nitrogenase gene amplicons from global marine surface waters are dominated by genes of non-cyanobacteria. *PLoS ONE* 6:e19223
- Fong AA, Karl DM, Lukas R, Letelier RM, Zehr JP, Church MJ (2008) Nitrogen fixation in an anticyclonic eddy in the oligotrophic North Pacific Ocean. *ISME J* 2:663–676
- Garcia N, Raimbault P, Sandroni V (2007) Seasonal nitrogen fixation and primary production in the Southwest Pacific: nanoplankton diazotrophy and transfer of nitrogen to picoplankton organisms. *Mar Ecol Prog Ser* 343:25–33
- Goebel NL, Edwards CA, Carter BJ, Achilles KM, Zehr JP (2008) Growth and carbon content of three different-sized diazotrophic cyanobacteria observed in the subtropical North Pacific. *J Phycol* 44:1212–1220
- Goebel NL, Turk KA, Achilles KM, Paerl R and others (2010) Abundance and distribution of major groups of diazotrophic cyanobacteria and their potential contribution to N₂ fixation in the tropical Atlantic Ocean. *Environ Microbiol* 12:3272–3289
- Hagino K, Onuma R, Kawachi M, Horiguchi T (2013) Discovery of an endosymbiotic nitrogen-fixing cyanobacterium UCYN-A in *Braarudosphaera bigelowii* (Prymnesiophyceae). *PLoS ONE* 8:e81749
- Halm H, Lam P, Ferdelman TG, Lavik G and others (2012) Heterotrophic organisms dominate nitrogen fixation in the South Pacific Gyre. *ISME J* 6:1238–1249
- Hashimoto R, Watai H, Miyahara K, Sako Y (2016) Title spatial and temporal variability of unicellular diazotrophic cyanobacteria in the eastern Seto Inland Sea. *Fish Sci* 82: 459–471
- Hewson I, Moisander PH, Morrison AE, Zehr JP (2007) Diazotrophic bacterioplankton in a coral reef lagoon: phylo-

- geny, diel nitrogenase expression and response to phosphate enrichment. *ISME J* 1:78–91
- Huang Y, Niu B, Gao Y, Fu L, Li W (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26:680–682
- Johnston AWB, Li Y, Ogilvie L (2005) Metagenomic marine nitrogen fixation: Feast or famine? *Trends Microbiol* 13: 416–420
- Karl D, Letelier R, Tupas L, Dore J, Christian J, Hebel D (1997) The role of nitrogen fixation in biogeochemical cycling in the subtropical North Pacific Ocean. *Nature* 388:533–538
- Kashtan N, Roggensack SE, Rodrigue S, Thompson JW and others (2014) Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* 344:416–420
- Knapp AN (2012) The sensitivity of marine N₂ fixation to dissolved inorganic nitrogen. *Front Microbiol* 3:374
- Kong L, Jing H, Kataoka T, Sun J, Liu H (2011) Phylogenetic diversity and spatio-temporal distribution of nitrogenase genes (*nifH*) in the northern South China Sea. *Aquat Microb Ecol* 65:15–27
- Krupke A, Musat N, LaRoche J, Mohr W and others (2013) In situ identification and N₂ and C fixation rates of uncultivated cyanobacteria populations. *Syst Appl Microbiol* 36: 259–271
- Krupke A, Lavik G, Halm H, Fuchs BM, Amann RI, Kuypers MMM (2014) Distribution of a consortium between unicellular algae and the N₂ fixing cyanobacterium UCYN-A in the North Atlantic Ocean. *Environ Microbiol* 16: 3153–3167
- Krupke A, Mohr W, Laroche J, Fuchs BM, Amann RI, Kuypers MMM (2015) The effect of nutrients on carbon and nitrogen fixation by the UCYN-A-haptophyte symbiosis. *ISME J* 9:1635–1647
- Langlois RJ, LaRoche J, Raab PA (2005) Diazotrophic diversity and distribution in the tropical and subtropical Atlantic Ocean. *Appl Environ Microbiol* 71:7910–7919
- Langlois RJ, Hümmer D, LaRoche J (2008) Abundances and distributions of the dominant *nifH* phylotypes in the northern Atlantic Ocean. *Appl Environ Microbiol* 74:1922–1931
- Langlois RJ, Mills MM, Ridame C, Croot P, LaRoche J (2012) Diazotrophic bacteria respond to Saharan dust additions. *Mar Ecol Prog Ser* 470:1–14
- Le Moal M, Biegala IC (2009) Diazotrophic unicellular cyanobacteria in the northwestern Mediterranean Sea: a seasonal cycle. *Limnol Oceanogr* 54:845–855
- Le Moal M, Collin H, Biegala IC (2011) Intriguing diversity among diazotrophic picoplankton along a Mediterranean transect: a dominance of rhizobia. *Biogeosciences* 8: 827–840
- Letunic I, Bork P (2011) Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucl Acids Res* 39:475–478
- Ludwig W, Strunk O, Westram R, Richter L and others (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* 32:1363–1371
- Luo YW, Doney SC, Anderson LA, Benavides M and others (2012) Database of diazotrophs in global ocean: abundance, biomass and nitrogen fixation rates. *Earth Syst Sci Data* 4:47–73
- Man-Aharonovich D, Kress N, Zeev EB, Berman-Frank I, Béjà O (2007) Molecular ecology of *nifH* genes and transcripts in the eastern Mediterranean Sea. *Environ Microbiol* 9:2354–2363
- Mazard SL, Fuller NJ, Orcutt KM, Scanlan DJ, Bridle O (2004) PCR analysis of the distribution of unicellular cyanobacterial diazotrophs in the Arabian Sea. *Appl Environ Microbiol* 70:7355–7364
- Messer LF, Doubell M, Jeffries TC, Brown MV, Seymour JR (2015) Prokaryotic and diazotrophic population dynamics within a large oligotrophic inverse estuary. *Aquat Microb Ecol* 74:1–15
- Messer LF, Mahaffey C, Robinson CM, Jeffries TC and others (2016) High levels of heterogeneity in diazotroph diversity and activity within a putative hotspot for marine nitrogen fixation. *ISME J* 10:1499–1513
- Moisander PH, Beinart RA, Hewson I, White AE and others (2010) Unicellular cyanobacterial distributions broaden the oceanic N₂ fixation domain. *Science* 327:1512–1514
- Moisander PH, Zhang R, Boyle EA, Hewson I, Montoya JP, Zehr JP (2012) Analogous nutrient limitations in unicellular diazotrophs and *Prochlorococcus* in the South Pacific Ocean. *ISME J* 6:733–744
- Montoya JP, Holl CM, Zehr JP, Hansen A, Villareal TA, Capone DG (2004) High rates of N₂ fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean. *Nature* 430:1027–1032
- Moore CM, Mills MM, Arrigo KR, Berman-Frank I and others (2013) Processes and patterns of oceanic nutrient limitation. *Nat Geosci* 6:701–710
- Musat N, Stryhanyuk H, Bombach P, Adrian L, Audinot JN, Richnow HH (2014) The effect of FISH and CARD-FISH on the isotopic composition of ¹³C- and ¹⁵N-labeled *Pseudomonas putida* cells measured by nanoSIMS. *Syst Appl Microbiol* 37:267–276
- Ratten JM, LaRoche J, Desai DK, Shelley RU and others (2015) Sources of iron and phosphate affect the distribution of diazotrophs in the North Atlantic. *Deep-Sea Res II* 116:332–341
- Rees AP, Turk-Kubo KA, Al-Moosawi L, Alliouane S, Gazeau F, Hogan ME, Zehr JP (2016) Ocean acidification impacts on nitrogen fixation in the coastal western Mediterranean Sea. *Estuar Coast Shelf Sci* 2:1–13
- Riou V, Fonseca-Batista D, Roukaerts A, Biegala IC and others (2016) Importance of N₂-fixation on the productivity at the northwestern Azores Current/Front System, and the abundance of diazotrophic unicellular cyanobacteria. *PLoS ONE* 11:e0150827
- Ryther JH, Dunstan WM (1971) Nitrogen, phosphorus, and eutrophication in the coastal marine environment. *Science* 171:1008–1013
- Scavotto RE, Dziallas C, Bentzon-Tilia M, Riemann L, Moisander PH (2015) Nitrogen-fixing bacteria associated with copepods in coastal waters of the North Atlantic Ocean. *Environ Microbiol* 17:3754–3765
- Shiozaki T, Chen YLL, Lin YH, Taniuchi Y, Sheu DS, Furuya K, Chen HY (2014) Seasonal variations of unicellular diazotroph groups A and B, and *Trichodesmium* in the northern South China Sea and neighboring upstream Kuroshio Current. *Cont Shelf Res* 80:20–31
- Shiozaki T, Nagata T, Ijichi M, Furuya K (2015) Nitrogen fixation and the diazotroph community in the temperate coastal region of the northwestern North Pacific. *Biogeosciences* 12:4751–4764
- Short SM, Zehr JP (2007) Nitrogenase gene expression in the Chesapeake Bay Estuary. *Environ Microbiol* 9: 1591–1596
- Sohm JA, Webb EA, Capone DG (2011) Emerging patterns of marine nitrogen fixation. *Nat Rev Microbiol* 9:499–508

- Sohm JA, Ahlgren NA, Thomson ZJ, Williams C and others (2016) Co-occurring *Synechococcus* ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. *ISME J* 10:333–345
- Subramaniam A, Yager PL, Carpenter EJ, Mahaffey C and others (2008) Amazon River enhances diazotrophy and carbon sequestration in the tropical North Atlantic Ocean. *Proc Natl Acad Sci USA* 105:10460–10465
- Thompson AW, Foster RA, Krupke A, Carter BJ and others (2012) Unicellular cyanobacterium symbiotic with a single-celled eukaryotic alga. *Science* 337: 1546–1550
- Thompson A, Carter BJ, Turk-Kubo K, Malfatti F, Azam F, Zehr JP (2014) Genetic diversity of the unicellular nitrogen-fixing cyanobacteria UCYN-A and its prymnesiophyte host. *Environ Microbiol* 16:3238–3249
- Tripp HJ, Bench SR, Turk KA, Foster RA and others (2010) Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium. *Nature* 464:90–94
- Turk KA, Rees AP, Zehr JP, Pereira N and others (2011) Nitrogen fixation and nitrogenase (*nifH*) expression in tropical waters of the eastern North Atlantic. *ISME J* 5: 1201–1212
- Turk-Kubo KA, Achilles KM, Serros TRC, Ochiai M, Montoya JP, Zehr JP (2012) Nitrogenase (*nifH*) gene expression in diazotrophic cyanobacteria in the tropical North Atlantic in response to nutrient amendments. *Front Microbiol* 3:386
- Turk-Kubo KA, Frank IE, Hogan ME, Desnues A, Bonnet S, Zehr JP (2015) Diazotroph community succession during the VAHINE mesocosm experiment (New Caledonia lagoon). *Biogeosciences* 12:7435–7452
- Villareal TA (1990) Laboratory culture and preliminary characterization of the nitrogen-fixing *Rhizosolenia-Richelia* symbiosis. *Mar Ecol* 11:117–132
- Villareal TA (1991) Nitrogen-fixation by the cyanobacterial symbiont of the diatom genus *Hemiaulus*. *Mar Ecol Prog Ser* 76:201–204
- Villareal TA (1994) Widespread occurrence of the *Hemiaulus*-cyanobacterial symbiosis in the southwest North Atlantic Ocean. *Bull Mar Sci* 54:1–7
- Villareal TA, Carpenter EJ (1989) Nitrogen fixation, suspension characteristics, and chemical composition of *Rhizosolenia* mats in the central North Pacific gyre. *Biol Oceanogr* 6:327–346
- Ward BA, Dutkiewicz S, Moore CM, Follows MJ (2013) Iron, phosphorus, and nitrogen supply ratios define the biogeography of nitrogen fixation. *Limnol Oceanogr* 58: 2059–2075
- Webb EA, Ehrenreich IM, Brown SL, Valois FW, Waterbury JB (2009) Phenotypic and genotypic characterization of multiple strains of the diazotrophic cyanobacterium, *Crocospaera watsonii*, isolated from the open ocean. *Environ Microbiol* 11:338–348
- Xiao P, Jiang Y, Liu Y, Tan W, Li W, Li R (2015) Re-evaluation of the diversity and distribution of diazotrophs in the South China Sea by pyrosequencing the *nifH* gene. *Mar Freshw Res* 66:681–691
- Zehr JP (2015) How single cells work together. *Science* 349: 1163–1164
- Zehr JP, Mellon MT, Zani S, York N (1998) New nitrogen-fixing microorganisms detected in oligotrophic oceans by amplification of nitrogenase (*nifH*) genes. *Appl Environ Microbiol* 64:3444–3450
- Zehr JP, Waterbury JB, Turner PJ, Montoya JP and others (2001) Unicellular cyanobacteria fix N₂ in the subtropical North Pacific Ocean. *Nature* 412:635–638
- Zehr JP, Montoya JP, Jenkins BD, Hewson I and others (2007) Experiments linking nitrogenase gene expression to nitrogen fixation in the North Pacific subtropical gyre. *Limnol Oceanogr* 52:169–183
- Zehr JP, Bench SR, Carter BJ, Hewson I and others (2008) Globally distributed uncultivated oceanic N₂-fixing cyanobacteria lack oxygenic photosystem II. *Science* 322:1110–1112
- Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30:614–620

Editorial responsibility: Eva Lindström,
Uppsala, Sweden

Submitted: April 27, 2016; Accepted: July 16, 2016
Proofs received from author(s): September 5, 2016