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 (N\textsubscript{2}) fixation is an enzyme-mediated process that is found among diverse lineages of Bacteria and Archaea, providing biologically available ammonium (NH\textsubscript{4}\textsuperscript{+}) from dinitrogen gas. N\textsubscript{2}-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 (CO\textsubscript{2}) and its export out of the photic zone.

© The authors 2016. Open Access under Creative Commons by Attribution Licence. Use, distribution and reproduction are unrestricted. Authors and original publication must be credited.

Publisher: Inter-Research · www.int-res.com
For many years, the paradigm was that N₂ 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 *Crocosphaera watsonii* (UCYN-B), contribute significantly to global N₂ 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, Song et al. 2008). UCYN-A *nifH* genes and *nifH* gene transcripts, a proxy for active N₂ 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 N₂ 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 N₂-fixing cyanobacteria. Subsequently, sequencing of the UCYN-A genome revealed an extremely unusual phototrophic microorganism that lacked the genes for oxygenic photosynthesis, CO₂ 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 N₂ (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 N₂ FIXATION RATES, GLOBAL DISTRIBUTION AND ABUNDANCES**

Since UCYN-A has not yet been cultured and it frequently co-occurs with other diazotrophs, studying specific N₂ 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 N₂ fixation rates (Montoya et al. 2004, Garcia et al. 2007, Biegala & Rainhaul 2008, Bonnet et al. 2009, Church et al. 2009, Benavides et al. 2011). The importance of unicellular cyanobacteria has been inferred from regions where N₂ 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 N₂ 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 N₂ 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 N₂ 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 ¹³C and ¹⁵N 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 N₂ 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 N₂ 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 N₂ fixation. In the tropical oligotrophic ocean, abundances of close to 10⁷ 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 N₂ 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 undersampled 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 N₂ fixation.
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 N₂ 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 N₂ fixation in N-replete waters (Knapp 2012) and lack of inhibition of N₂ fixation activity in nitrate amended cultures of the unicellular diazotrophic cyanobacterium Crocosphaera (Dekaezemacker & Bonnet 2011) indicate that the regulation of N₂ 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 N₂ 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 N₂ 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 *Crococapsa* (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 *Cyanothecaceae*. 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
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; Table S2 in Supplement 2; Messer et al. 2015: 658943 reads; Tilia et al. 2015: 25841 reads; Table S2 in Supplement 2; Messer et al. 2016: 18331 reads; Bentzon-Tilia 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,
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 \textit{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 \textit{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, \textit{nifH} abundance and transcription may have been underestimated 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 $2 \times 10^3$ \textit{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 $7.6 \times 10^5$ \textit{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 N$_2$ fixation. Unfortunately, due to the short \textit{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 \textit{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 \textit{nifH} sequences and the discovery of closely related sublineages, it will be important to re-evaluate the amplification specificity and cover-

<table>
<thead>
<tr>
<th>qPCR assay name (reference sequence if available)</th>
<th>Number of total mismatches of primers and probes to the OTU representative of each sublineage</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>OTU0002 (gi 213578830)</td>
<td>UCYN-A1 5 0 0 0</td>
<td>Bentzon-Tilia et al. (2015)</td>
</tr>
<tr>
<td>Group A (AF059642)</td>
<td>UCYN-A2 2 5 4 4</td>
<td>Church et al. (2005a)</td>
</tr>
<tr>
<td>UCYN-A (HM210392)</td>
<td>UCYN-A3 1 1 0 2</td>
<td>Halm et al. (2012)</td>
</tr>
<tr>
<td>UCYN-A-nifH-2</td>
<td>UCYN-A4 0 0 2 1</td>
<td>Hashimoto et al. (2016)</td>
</tr>
<tr>
<td>UCYN-A (HQ456651)</td>
<td></td>
<td>Kong et al. (2011)*</td>
</tr>
<tr>
<td>Group A (AF059627)</td>
<td></td>
<td>Langlois et al. (2008)</td>
</tr>
<tr>
<td>Group A (AB665496)</td>
<td></td>
<td>Shiozaki et al. (2014)</td>
</tr>
<tr>
<td>CBga (CB4713A1)</td>
<td></td>
<td>Short &amp; Zehr (2007)</td>
</tr>
<tr>
<td>UCYN-A2</td>
<td></td>
<td>Thompson et al. (2014)</td>
</tr>
</tbody>
</table>

*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 *Crocosphaera* 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 phylotypes. 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 pynmensiophyte 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-
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.

<table>
<thead>
<tr>
<th>UCYN-A cell diameter (µm)</th>
<th>N</th>
<th>Eukaryotic cell diameter (µm)</th>
<th>N</th>
<th>UCYN-A probe</th>
<th>Eukaryotic probe</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.62 ± 0.16</td>
<td>19</td>
<td>1.36 ± 0.24</td>
<td>10</td>
<td>UCYN-A732</td>
<td>Not used</td>
<td>Thompson et al. (2012)</td>
</tr>
<tr>
<td>1.36 ± 0.45</td>
<td>44</td>
<td>3.58 ± 1.42</td>
<td>8</td>
<td>UCYN-A732, UCYN-A159</td>
<td>Not used</td>
<td>Krupke et al. (2013)</td>
</tr>
<tr>
<td>1.16 ± 0.24</td>
<td>52</td>
<td>2.26 ± 0.75</td>
<td>52</td>
<td>UCYN-A732</td>
<td>PRYM02</td>
<td>Krupke et al. (2014)</td>
</tr>
<tr>
<td>0.83 ± 0.15</td>
<td>44</td>
<td>1.66 ± 0.23</td>
<td>44</td>
<td>UCYN-A732</td>
<td>PRYM02</td>
<td>Krupke et al. (2015)</td>
</tr>
<tr>
<td>na</td>
<td>2030</td>
<td>2.5</td>
<td>17</td>
<td>UCYN-A732</td>
<td>UPRYM-69 (UCYN-A1 partner)</td>
<td>Cabello et al. (2016)</td>
</tr>
<tr>
<td>na</td>
<td>4-5</td>
<td>17</td>
<td>UCYN-A732</td>
<td>UBRADO-69 (UCYN-A2 partner)</td>
<td>Cabello et al. (2016)</td>
<td></td>
</tr>
</tbody>
</table>
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 N₂ 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


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 nontypical nitrogen-fixing cyanobacteria. PLoS ONE 6:e19223


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


Messer LF, Mahaffey C, Robinson CM, Jeffries TC and others (2016) High levels of heterogeneity in diazotrophic diversity and activity define a putative hotspot for marine nitrogen fixation. ISME J 10:1499–1513


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


