Genetic diversity of the unicellular nitrogen-fixing cyanobacteria UCYN-A and its prymnesiophyte host

Anne Thompson [1 *], Brandon J. Carter [1], Kendra Turk-Kubo [1], Francesca Malfatti [2 ‡], Farooq Azam [2], Jonathan P. Zehr [1]

[1] Ocean Sciences Department, University of California - Santa Cruz
1156 High Street, CA 95064, USA.

[2] Scripps Institution of Oceanography, University of California - San Diego
San Diego, CA.

[*] Corresponding author. Current address: BD Biosciences Advanced Cytometry Group, 12730 28th Ave NE, Seattle, WA 98125. (206) 364-3400. anne.w.thompson@gmail.com

[‡] Current address: National Institute of Oceanography and Experimental Geophysics, Santa Croce, Trieste, Italy.

Running title: UCYN-A genetic diversity

Summary

Symbiotic interactions between nitrogen-fixing prokaryotes and photosynthetic eukaryotes are an integral part of biological nitrogen fixation at a global scale. One of these partnerships involves the cyanobacterium UCYN-A, which has been found in partnership with an uncultivated unicellular prymnesiophyte alga in open-ocean and coastal environments. Phylogenetic analysis of the UCYN-A nitrogenase gene (*nifH*) showed that the UCYN-A lineage is represented by three distinct clades, referred to herein as UCYN-A1, UCYN-A2 and UCYN-A3, which appear to have overlapping and distinct geographic distributions. The relevance of UCYN-A's genetic diversity to its symbiosis and ecology was explored through combining flow cytometric cell sorting and molecular techniques to determine the host identity, *nifH* expression patterns, and cell size of one newly-discovered clade, UCYN-A2, at a coastal site. UCYN-A2 *nifH* expression peaked during daylight hours, which is consistent with expression patterns of the UCYN-A1 clade in the open ocean. However, the cell size of the UCYN-A2 host was significantly larger than UCYN-A1 and host, suggesting adaptation to different environmental conditions. Like the UCYN-A1 host, the UCYN-A2 host was closely related to the genus *Braarudosphaera*, however the UCYN-A1 and UCYN-A2 host rRNA sequences clustered into two distinct clades suggesting co-evolution of symbiont and host.

Introduction

Symbioses between nitrogen (N_2)-fixing prokaryotes (diazotrophs) and photosynthetic eukaryotes play an important role in global biological nitrogen fixation (BNF) (Carpenter and Foster 2003; Houlton et al. 2008; Foster et al. 2011; Karl et al. 2012). In the oceans, these symbioses usually involve diazotrophic cyanobacteria and photosynthetic eukaryotic plankton and the partnerships range from obligate to facultative (Carpenter and Foster 2003; Lesser et al. 2004; Foster et al. 2006; Foster et al. 2011; Hilton et al. 2013).

Of the symbiotic marine diazotrophic cyanobacteria, the unicellular UCYN-A is recognized for its global distribution (Goebel et al. 2010; Moisander et al. 2010; Zehr and Kudela 2011) and significant contributions to local N_2-fixation (Montoya et al. 2004). The genome of UCYN-A is extremely streamlined and lacks genes encoding the photosystem II complex (PSII), carbon fixation pathways, and various other pathways including the TCA cycle (Zehr et al. 2008; Tripp et al. 2010). A second UCYN-A genome was recently sequenced and is missing the same pathways, suggesting that a streamlined genome and the absence of essential metabolic pathways is a defining feature of the UCYN-A lineage (Bombar et al. *Submitted*). It appears that the streamlined genome of UCYN-A is a result of genome degradation through obligate symbiosis with a unicellular eukaryotic host. The eukaryotic host is from a lineage of uncultivated prymnesiophytes related to *Braarudosphaera bigelowii*, a calcareous phytoplankton, which was observed in symbiosis with UCYN-A in the open ocean (Thompson et al. 2012) and in the coastal waters of Japan (Hagino et al. 2013). Incubation of seawater with labeled nutrients demonstrated N transfer from the cyanobacterium to the host, and C transfer from host to cyanobacterium (Thompson et al. 2012; Krupke et al. 2013), suggesting that the

basis of the symbiosis is nutrient exchange, specifically fixed N and fixed C, between the partner species.

While it is recognized that UCYN-A is likely to be one of the major marine diazotrophs, the genetic diversity of UCYN-A is not well known. For the very abundant marine non-diazotrophic cyanobacteria, *Prochlorococcus* and *Synechococcus,* and the less abundant diazotrophic cyanobacteria *Crocosphaera,* genetic diversity has clear relevance to ecological function (Kettler et al. 2007; Webb et al. 2009; Coleman and Chisholm 2010; Malmstrom et al. 2010; Bench et al. 2013). In these cases, genetically-distinct clades (or, ecotypes) exhibit phenotypic traits that reflect local environmental conditions including light availability, nutrient composition, and temperature. Understanding the ecological function of the different ecotypes has been critical to interpreting measurements of ecotype abundance and the content of environmental sequence databases for these microbial groups. However, it has been difficult to identify analogous ecotypes in UCYN-A due to its uncultivated status, relatively low abundances, and poor to nonexistent representation in large metagenomic databases.

In this study, the genetic diversity of UCYN-A was assessed from phylogenetic analysis of a large database of nitrogenase (*nifH*) sequences from marine systems, which revealed three distinct clades of UCYN-A. One of the clades that emerged from the phylogenetic analysis is well-represented in populations collected from the Scripps Institute of Oceanography (SIO) Pier (San Diego, California, USA) and is the focus of this study.

Results and Discussion

Divergence of UCYN-A into three distinct clades

Phylogenetic analysis of UCYN-A nitrogenase (*nifH*) gene sequences revealed divergence of the UCYN-A lineage into multiple clades (Figure 1A). The clades were designated UCYN-A1, UCYN-A2, and UCYN-A3, with clade UCYN-A1 containing the original UCYN-A genome (Tripp et al. 2010). Variation between the sequences occurred predominantly in the third base pair position of each codon and clades could not be differentiated by *nifH* amino acid sequences (Figure 1B). Functional restraint on sequence variance in the nitrogenase protein likely explains the lack of divergence in the amino acid sequences. However, divergence of *nifH* at the nucleotide level may indicate that each UCYN-A clade was subject to different evolutionary pressures after divergence from a common ancestor. Indeed, this is consistent with the results of a comparative genomic study between UCYN-A1 and UCYN-A2, which shows that though over 96% of UCYN-A1 genes are present in UCYN-A2, the shared genes are on average only 86% similar at the amino acid level (Bombar et al. *Submitted*).

To determine if the UCYN-A clades occupied overlapping or distinct oceanic habitats, the geographic distribution of UCYN-A *nifH* sequences submitted to GenBank from clone-based studies was examined (Figure 2). UCYN-A1 *nifH* sequences have been recovered from many of the major ocean basins, however all clades were widely distributed (Figure 2). UCYN-A2 *nifH* sequences were present where large numbers of UCYN-A1 had been reported from clone libraries, with the exception of the Gulf of Catalina (this study), Heron Reef (Hewson et al. 2007), and the Eastern Mediterranean Sea (Man-Aharonovich et al. 2007). UCYN-A3 *nifH* sequences have also been reported in regions where other UCYN-A clades were present, such as the Cape Verde Islands in the Eastern Atlantic, but also in the South Pacific Gyre where cyanobacterial diazotrophs are scarce (Halm et al. 2012; Turk-Kubo et al. 2013). Several sequences from the Eastern Atlantic, where diazotroph diversity is known to be rich (Turk et al.

2011), did not cluster clearly with any clade (Figure 1 and 2D), indicating that additional clades may exist.

The discovery of additional UCYN-A clades has implications for estimates of the contribution of UCYN-A to oceanic N_2 fixation. The most frequently used quantitative assay for UCYN-A was designed from Station ALOHA UCYN-A1 sequences (Church et al. 2005) and has been used to assay UCYN-A abundances in other ocean basins (Goebel et al. 2010; Moisander et al. 2010). However, this assay (now referred to as the UCYN-A1 *nifH* assay) does not detect UCYN-A2 or UCYN-A3 due to mismatches at several positions (Supporting Information - Table S1). As a result, the true population size of UCYN-A and their N_2 fixation rates has been underestimated.

We investigated the characteristics of one of the newly recognized UCYN-A clades, UCYN-A2, in the Gulf of Santa Catalina, a region where UCYN-A1 has not previously been reported (Figure 2). Subsequent sampling indicated that UCYN-A2 was abundant at the Scripps Institute of Oceanography (SIO) Pier in the Gulf of Santa Catalina (Supporting Information - Table S2) and clone libraries indicated that UCYN-A *nifH* sequences recovered from these samples were exclusively UCYN-A2.

Diel nitrogenase expression by UCYN-A2

The expression of nitrogenase during the day is a feature of UCYN-A1 that sets it apart from most other unicellular N_2-fixing cyanobacteria such as *Crocosphaera* and *Cyanothece*, that fix N_2 at night to protect nitrogenase from oxygen (Schneegurt et al. 1997; Church et al. 2005). Daytime N_2 fixation by UCYN-A1 is likely possible because genes responsible for oxygen generation via photosynthesis are missing (Zehr et al. 2008) and UCYN-A may be physically

isolated from its photo-oxygenic host by membrane(s) boundaries, as has been shown for a UCYN-A isolated in coastal Japan (Hagino et al. 2013). To determine whether UCYN-A2 is similar to UCYN-A1 in diel patterns of N_2 fixation, UCYN-A2 *nifH* gene expression was measured over three consecutive diel cycles at the SIO Pier. Maximum *nifH* expression occurred during the morning between 0600 and 1200 hours, which is consistent with the timing of *nifH* expression by UCYN-A1 at Station ALOHA (Figure 3) and the lack of oxygen-evolving pathways in the genome of UCYN-A2 (Bombar et al. *Submitted*). When comparing the peak-to-peak amplitude of gene expression, UCYN-A2 *nifH* expression at SIO spanned 4 orders of magnitude while UCYN-A1 *nifH* expression at the open ocean station ALOHA spanned only 2 orders of magnitude (Figure 3). This difference may be due to different sampling depths between the two stations. SIO Pier samples were collected at the surface while Station ALOHA samples were collected at 25 m depth (Church et al. 2005). At depth, diel fluctuations in light availability are dampened due to the attenuation of light in water with depth. This observation may suggest that UCYN-A is sensitive to light availability despite the absence of major photosynthetic machinery such as the Photosystem II complex. Whether these transcriptional responses are mediated through the host, or autonomously, remains unclear.

UCYN-A2 host identification

Genetic specificity between host and symbiont pairings can indicate co-evolution between symbiotic partners. To determine if the UCYN-A clades associate with different hosts, cell sorting and qPCR were applied to untreated samples from the SIO Pier in order to identify the UCYN-A2 host. Distinct phytoplankton populations were analyzed by flow cytometry and isolated by sorting from five sort gates in October 2012 and June 2013. Cells from sort gates A

and B were positive for UCYN-A2 *nifH*, while gates C, D, and E were negative (Figure 4). The highest proportion of *nifH* gene copies amplified per sorted cell was from sort gate A. Thus, sort gate A was targeted for subsequent sorts, microscopy, and screenings of single phytoplankton cells for UCYN-A2 host identification.

UCYN-A2 *nifH* was amplified from only 9.5% of single phytoplankton cells (10 cells) from gate A, suggesting that gate A contained a mix of cell types, only some of which contained UCYN-A. Microscopic analysis of sorted cells from gate A revealed diatoms, dinoflagellates, unidentified flagellated and non-flagellated cells, however, the host could not be positively identified among these cells (Supporting Information - Figure S1). Together microscopy and flow cytometry indicated that the UCYN-A host contained chlorophyll but not phycoerythrin and ranged from 7 – 10 μm in diameter (Figure 4, Supporting Information - Figure S1). This size is significantly larger than the estimates of UCYN-A1 host diameter at 1 – 3 μm (Thompson et al. 2012; Krupke et al. 2013). Differences in size may indicate different nutrient acquisition strategies and requirements of the host associated with each lineage of UCYN-A. The importance of cell size to nutrient acquisition in phytoplankton has been tested explicitly for diatoms and iron where cell size was inversely correlated with iron uptake rates (Sunda and Huntsman 1995; Sunda and Huntsman 1997). Similar effects are observed for a range of phytoplankton species, including the prymnesiophyte *E. huxleyi* (Sunda and Hardison 2010).

Ten UCYN-A2-positive single phytoplankton cells from gate A were subjected to nested PCR using universal 18S rRNA gene primers in order to identify the host. Seven of the cells yielded prymnesiophyte sequences (GenBank accession numbers KF771248-KF771254) that were 100% identical (6 cells) and 99.1% (1 cell) to *Braarudosphaera bigelowii* strains Yatsushiro 1 (GenBank accession number AB478414), Furue 15 (GenBank accession number

175 AB478413), and TP05-6-a (GenBank accession number AB058358) (Figure 5). The three
176 remaining nested PCRs yielded only non-marine Chinese white pine (*Pinus armandii*) 18S rRNA
177 gene sequences that were also present in negative controls in this and previous studies
178 (Thompson et al. 2012). The UCYN-A2 host 18S rRNA gene is closely related to, but only has
179 only 98% nucleotide similarity across the amplified region to that of the UCYN-A1
180 prymnesiophyte host (Gen Bank accession number JX291893). This result confirms a recent
181 study that amplified UCYN-A 16S rRNA genes from an isolated specimen of Intermediate
182 Form-B *B. bigelowii* (Hagino et al. 2013).
183 The prymnesiophytes are abundant, ecologically relevant, and diverse eukaryotic
184 phytoplankton (Jardillier et al. 2010) and *Braarudosphaera* is no exception. Extant
185 *Braarudosphaera* have been recovered from diverse environments including coastal Japan
186 (Takano et al. 2006; Hagino et al. 2009), the Bering Sea (Konno et al. 2007), the Sargasso Sea
187 (Gaarder 1954; Hulburt 1962) and the Mediterranean Sea (Borsetti and Cati 1972;
188 Knappertsbusch 1993) (see Konno et al. (2007) for additional references and distribution map).
189 *B. bigelowii* is known to exist as at least five different genotypes and 18S rRNA gene sequences
190 have been recovered from calcareous pentalith-forming *B. bigelowii* in coastal Japanese waters
191 (Takano et al. 2006; Hagino et al. 2009). Thus, it is particularly intriguing that the coastally-
192 derived UCYN-A2 host sequence from this study (SIO Pier) clusters with *B. bigelowii* 18S
193 rRNA gene sequences from coastal Japan (Takano et al. 2006; Hagino et al. 2009; Hagino et al.
194 2013) while the UCYN-A1 host sequence clusters with sequences derived from the open-ocean
195 (Figure 5). There is now evidence that both *B. bigelowii* Intermediate form B (Hagino et al.
196 2013) and an open-ocean close relative of *B. bigelowii* (Thompson et al. 2012) harbor UCYN-A.
197 This study demonstrates that at least for UCYN-A1 and UCYN-A2, the host-symbiont

198 relationships are formed between genetically-distinct UCYN-A and *B. bigelowii* lineages.
199 However, it is still unknown whether other types of *B. bigelowii* (Takano et al. 2006; Hagino et
200 al. 2009) form symbioses with UCYN-A. If symbionts are always present, it is possible that the
201 genetic and morphological diversity of *B. bigelowii* corresponds to the diversity of UCYN-A,
202 and that there has been co-evolution of *B. bigelowii* and UCYN-A strains.
203 In previous studies, the UCYN-A1 host was not genetically identical to cells where the
204 presence of plates was confirmed, therefore the calcareous nature of the UCYN-A1 was
205 speculative (Thompson et al. 2012). However, the 100% genetic match of the UCYN-A2 host
206 sequence to *B. bigelowii* cells known to carry calcareous plates and a recent microscopic
207 observation of UCYN-A within a pentalith-carrying cell demonstrate unequivocally that the
208 UCYN-A host is capable of forming calcareous plates (Hagino et al. 2013). However,
209 picoeukaryotes with calcareous plates were not apparent following microscopic observation of
210 UCYN-A2-positive phytoplankton from sorted samples in this study (Supporting Information -
211 Figure S1). One possible explanation is that *B. bigelowii* passes through a non-calcifying life
212 stage during which this study was conducted (June and October). Indeed, other calcifying
213 prymnesiophytes carry out heteromorphic life cycles. *Emiliana huxleyi*, for example, exists in a
214 diploid calcifying form, a diploid non-calcifying form, and a haploid non-calcifying form. Work
215 on cultured isolates and naturally-occurring populations of *E. huxleyi* demonstrate that each life
216 stage exhibits distinct physiological and genetic characteristics such as light physiology and viral
217 susceptibility (Houdan et al. 2005; von Dassow et al. 2009; Frada et al. 2012). Similar
218 physiological changes may be part of the UCYN-A host life cycle.
219 The patterns of *B. bigelowii* calcification and the impacts of this host process on the UCYN-
220 A symbiont are relevant in the context of changing carbonate chemistry and rising atmospheric

carbon dioxide (CO_2) concentrations. For an organism such as *E.huxleyi*, the relationship between calcification and CO_2 concentration is not straightforward (Beaufort et al. 2011). The case may be similar for *B. bigelowii,* as this group is present in diverse ecosystems with varied chemical gradients (Takano et al. 2006; Hagino et al. 2009). If host calcification sensitivities and patterns influence UCYN-A N_2 fixation rates, there may be as of yet unexplored consequences of global climate change to the global nitrogen cycle (Doney et al. 2009).

UCYN-A2 and host abundance at the SIO Pier

The abundances of UCYN-A2 and its host were measured in two sets of samples taken at different times from the SIO Pier using qPCR targeting *nifH* and the 18S rRNA gene, respectively. A monthly sampling series revealed the highest abundances of UCYN-A2 between May and October 2011 (up to 1.9 $x10^5$ *nifH* copies per L) while the lowest abundances occurred between November and April 2011 (1.3 x 10^3 *nifH* copies per L) (Supporting Information - Table S2). An hourly sampling series from a 3-day period in October 2012 revealed stable abundances of UCYN-A2 and *B. bigelowii* despite tidal flux and shifts in temperature, chlorophyll *a* concentration, and salinity (Figure 6, Supporting Information - Figure S2). UCYN-A2 abundances averaged 1.36 x 10^5 *nifH* copies per L during the three-day period and host abundances remained constant with an average of 4.3 x 10^4 18S rRNA gene copies per L of seawater. Comparable ranges of UCYN-A1 abundances have been measured at Station ALOHA (Church et al. 2005). Their persistent abundance through physical changes in the environment may suggest that *B.bigelowii* abundance is more heavily controlled by nutrient availability (not measured) than light availability and temperature, however, the influence of the environment on *B. bigelowii* ecology is poorly understood.

The relative abundance of two symbiotic partners is an important aspect of the ecology and evolution of host-symbiont interactions. In the case of UCYN-A and its host, the relative abundances of the partners will aid in estimating contributions of UCYN-A to N_2 fixation taking into account host N requirements, nutrient and metabolite exchange between the two cells, and the mechanism of UCYN-A transfer between host generations. In this study, the ratio of UCYN-A2 to *B. bigelowii* abundance in bulk DNA samples ranged from 0.2 to 11.0 over three days, averaging 3.3 UCYN-A2 cells per host (derived from data in Figure 6, and assuming one copy of *nifH* in UCYN-A2 and one copy if the 18S rRNA gene in *B. bigelowii*). It must be noted that these ratios are estimates as qPCR is not reliable for absolute quantification of single gene copies, and *B. bigelowii* may have multiple copies of the 18S rRNA gene like many other picoeukaryotes (Prokopowich et al. 2003; Zhu et al. 2005). Previously, application of Halogenated *In Situ* Hybridization (HISH) and Fluorescence *In Situ* Hybridization (FISH) to seawater samples detected 1-2 UCYN-A1 per host (Thompson et al. 2012; Krupke et al. 2013). A microscopy study of *B. bigelowii* in coastal Japan waters showed that there were either one or two UCYN-A cells per host (Hagino et al. 2013). These independent lines of evidence suggest that at least 1-2 UCYN-A symbionts are associated with each host, however it remains unclear if the UCYN-A cells are sister cells and how their division might be synchronized with host cell division and life cycles.

Conclusions

The results of this study showed that one of the major lineages of marine diazotrophic cyanobacteria, UCYN-A, is represented by three genetically-distinct clades. Two of these clades, UCYN-A1 and UCYN-A2, were compared and were similar in their day-time expression of *nifH*

and symbiosis with a *B.bigelowii*-like prymnesiophyte. However, the study demonstrated that UCYN-A1 and UCYN-A2 associate with specific host genotypes suggesting coevolution of symbiont and host and the adaptation of each host-symbiont pair to a specific environmental niche. In some cases, these niches are physically distant and distinct in nutrient availability and oceanographic conditions, such as in the case of UCYN-A1 from Station ALOHA in the open ocean and UCYN-A2 from the coastal SIO Pier. However, in most of the environments that have been sampled, multiple clades of UCYN-A are present, thus specific host-symbiont pairs may occupy overlapping ecological niches as well. Future work will reveal if the genetic distinction between lineages of the UCYN-A symbiont and prymnesiophyte host have implications to patterns of biological nitrogen fixation in the marine environment.

Experimental Procedures

Sample and oceanographic data collection

Seawater samples were collected with a bucket from the end of the Scripps Institution of Oceanography (SIO) Ellen Browning Scripps Memorial Pier in La Jolla, CA. After collection, sample water was gently poured into a 10 L polypropylene bottle then placed in a dark cooler and transported to the lab (5 minutes). The first set of samples was collected monthly between September 2010 and May 2011, then they were shipped overnight to Santa Cruz, CA for filtration and DNA analysis. The second set of samples was collected October 1st - 4th 2012, May 8th 2013, and June 10 -13th 2013 and was processed immediately for flow cytometry, gene expression, and DNA analysis. Oceanographic data for chlorophyll concentration, temperature, and salinity were collected from the Southern California Coastal Ocean Observation System website (www.sccoos.org/) for the dates sampled.

Flow cytometry

Flow cytometry was performed using the Influx (Becton Dickinson, San Jose, CA, USA) high-speed cell sorter equipped with a small particle detector and a 488 nm laser (Sapphire Coherent, Coherent Inc., Santa Clara, CA, USA). In the course of the work presented here, two different cytometers were used, thus cytometry settings vary based on the instrument used for data collection and the experimental purpose of the measurements. For example, the 3 μm beads in Figure 4A and B are used for qualitative reference rather than calibration of the forward scatter (FSC) signal in each panel. BioSure (Grass Valley, CA, USA) Sheath Solution was used at 1X concentration with either 70 or 100 μm diameter nozzles. Data collection and sorting was triggered with light scattered in the forward direction (FSC). Sortware (Becton Dickinson) data collection software was used to identify populations of cells for gating and cell sorting. Beads of 0.75, 3, 10, and 25 μm diameter Polysciences Fluoresbrite ® YG Microspheres and/or 3.8 μm diameter Ultra Rainbow Fluorescent Particles (Spherotech, Warrington, PA, USA) were used for internal reference. Data analysis and figures were created in Flow Jo (TreeStar, Ashland, OR, USA).

Prior to cell analysis and sorting, seawater samples were either untreated (not filtered or frozen, no preservatives) or were agitated. Untreated seawater was used to study the intact UCYN-A2/host complex for host identification and characterization experiments. Agitated seawater samples were used for experiments to study UCYN-A2 when separated from the host. Agitation involved vacuum filtration (5 – 10 psi) onto 0.22 μm pore-size type GV 47 mm diameter filters (Millipore, Billerica, CA, USA), 100X concentration by re-suspension of the

312 filter in sterile-filtered seawater, brief vortex mixing, flash-freezing in liquid nitrogen, thawing at
313 room-temperature, and 1 minute of vortex mixing before analysis.
314
315 *Nucleic acid extraction and complementary DNA synthesis*
316 RNA and DNA samples were collected by vacuum filtration of replicate 500 mL volumes
317 of seawater at 10 psi pressure onto 47 mm, 0.22 μm pore-size, Supor filters (Pall Corporation,
318 Port Washington, NY, USA). Filters were placed in sterile 2 mL bead-beating tubes with sterile
319 glass beads (both DNA and RNA samples). RNA samples were amended with the lysis buffer
320 RLT (Qiagen, Valencia, CA, USA) and β-mercaptoethanol before storage. Samples were stored
321 at -80 °C until extraction. Care was taken to process RNA samples within 15 minutes of seawater
322 sampling.
323 DNA extractions were carried out using a modification of the Qiagen DNeasy Plant Kit
324 (Moisander et al. 2008). Samples were subjected to 2 minutes of bead-beating, three sequential
325 freeze-thaw cycles using liquid nitrogen and a 60 °C water bath, a one hour proteinase K
326 treatment, and column purification using the QIAcube automated extraction platform (Qiagen).
327 RNA extraction was performed using a modified version of the Qiagen RNeasy Plant
328 Mini Kit, which included an automated on-column DNase treatment step. The QuantiTect
329 Reverse Transcription Kit (Qiagen) was used for additional removal of genomic DNA and
330 synthesis of complementary DNA (cDNA) (Turk-Kubo et al. 2012).
331
332 *Quantitative PCR (qPCR) assay design and application*
333 Two Taqman® quantitative PCR (qPCR) assays, one targeting the UCYN-A2
334 nitrogenase (*nifH*) gene and the other targeting the *Braarudosphaera sp.* 18S rRNA gene

sequence present at the SIO Pier, were developed to quantify UCYN-A2 and host from DNA samples, whole sorted cells, and cDNA (UCYN-A2 only).

Nested PCR using degenerate *nifH* primers (Zehr and McReynolds 1989; Zani et al. 2000) was applied to the first set of DNA samples (DNA collected monthly). Sequences (GenBank accession numbers KF806604-KF806612) belonged exclusively to the UCYN-A2 clade and were used to inform design of a UCYN-A2-specific qPCR assay. Primers and probe sequences for the UCYN-A2 *nifH* gene assay (96 bp amplicon) are as follows: Forward, 5'-GGTTACAACAACGTTTTATGTGTTGA-3'; Reverse, 5'-ACCACGACCAGCACATCCA-3'; Probe, 5'-FAM-TCTGGTGGTCCTGAGCCCGGA-TAMRA-3'. Nested PCR with universal 18S rRNA gene primers (see below), were used to amplify the 18S rRNA gene of UCYN-A2 host cells (*Braarudosphaera bigelowii*), which were used to design the UCYN-A2 host qPCR assay. Primers and probe sequences sequences for the SIO *B. bigelowii* 18S rRNA gene assay (96 bp amplicon) are as follows: Forward, 5'-GGTTTTGCCGGTCTGCCGTT-3'; Reverse, 5'-ATCCGTCTCCGACACCCACTC-3'; Probe, 5'-FAM-CTGGTGCGAGCGTCCTTCCT-TAMRA-3'.

Each assay used TaqMan® Gene Expression MasterMix (Invitrogen, Carlsbad, CA, USA) at 1X concentration along with 0.4 μM forward and reverse primers and 0.2 μM probe. Both assays were initially incubated for 10 minutes at 95 °C to relax target DNA and data was collected at the end of each of 45 staged repeats of 15 seconds at 95 °C and 60 seconds at either 64 °C (UCYN-A2 *nifH* gene) or 60 °C (SIO *Braarudosphaera* 18S rRNA gene). Standards for each assay were generated using linear plasmids containing clones of PCR amplified gene from environmental samples containing either UCYNA-A2 *nifH* (359 bp) or *Braarudosphaera* sp. 18S rRNA (733bp).

Identification of target cell populations by flow cytometry and qPCR

The UCYN-A2-specific qPCR assay was used to screen flow-sorted events (individual phytoplankton cells) to identify populations enriched in UCYN-A2 (sorts of 50-500 events) or individual phytoplankton (1 event) that could contain both UCYN-A2 and its host organism. Cells were sorted into aliquots of 10 µl, 5 kDa-filtered, nuclease-free water then amended for "whole cell" qPCR as above. Assays were run with an Applied Biosystems 7500 Real Time PCR thermal cycler.

Confocal Microscopy

Once the seawater was sorted and the presence of *nifH* gene was confirmed by qPCR, the positive *nifH* gene fraction was imaged live in a chamber slide at 600X final magnification at the Laser Scanning Confocal Microscope A1R (Nikon Instrument Inc. Japan). The nucleic stain DAPI, was added to the seawater fraction at a final concentration of 10 µg mL^{-1}. The sample was imaged using the 405 nm (DAPI) laser, the 488 nm (phycoerythrin) laser and the 648 nm (chlorophyll) laser.

Nested PCR on single cells with universal 18S rRNA gene primers

Single events (individual phytoplankton cells) from a sort gate enriched in UCYN-A2 were sorted into individual wells of 96-well qPCR plates on October 4, 2012 and June 12, 2013. Plates were screened for UCYN-A2 using qPCR alongside standards and no template controls (NTC). Wells identified as positive for UCYN-A2 *nifH* were selected and the entire reaction

volume (25 µl) was used as template in a nested PCR series using universal 18S rRNA gene primers.

The nested PCR series consisted of two sequential reactions: the first was primed with external primers EukA/EukB (Medlin et al. 1988) and the second was primed with internal primers Euk555F/Univ1269R (López-García et al. 2003). EukA/EukB reactions were run in 100 µl volumes containing 75 µl of master mix added to the 25 µl volume of a UCYN-A2 positive qPCR. PCR using Euk555F/Univ1269R primers was then performed in 50 µl volumes with 2 µl of the EukA/EukB reaction as template. All PCR was performed with ExTaq DNA polymerase (Takara Bio, Inc., Otsu, Shiga, Japan) with 25 mM $MgCl_2$. Gel electrophoresis was used to visualize PCR product from the inner primers. Amplified DNA bands of approximately 700 bp were cut and purified from the gel. The pGEM-T Vector System (Promega, Madison, WI, USA) was used for cloning of 18S rRNA gene products and Sanger sequencing was performed at the UC Berkeley DNA Sequencing Center (Berkeley, CA). Trimming of vectors, removal of poor quality sequences, and sequence comparisons were performed using Sequencer version 5.1 (Gene Codes Corporation, Ann Arbor, MI, USA). GenBank accession numbers for the 7 UCYN-A2 host 18S rRNA gene sequences that were recovered are KF771248-KF771254.

Phylogenetic analysis

UCYN-A-like sequences were identified using an auto-curated publically available database (http://pmc.ucsc.edu/~wwwzehr/research/database/). This database was populated with all *nifH* sequences submitted to GenBank's nr/nt database as of September 6, 2013. Amino acid sequences were aligned using the Hidden Markov Model profile Fer4_NifF_fs (Pfam PF00142) in HMMer v3.1b1 (Finn et al., 2010), and nucleotide sequences were back-aligned to the aligned

amino acids. UCYN-A-like sequences were identified based on their amino acid similarity to the UCYN-A1 genome (Tripp et al. 2010) and were clustered at 97% nucleic acid identity, using the CD-HIT package (Li and Godzik 2006). Representative sequences of the resulting clusters were chosen for the construction of nucleic acid and amino acid trees (Figure 1A, B). Maximum likelihood trees of partial *nifH* nucleic acid and amino acid sequences were constructed in MEGA6.06 (Ludwig et al 2004). Evolutionary distances were inferred using the Tamura-Nei (307 positions) and JTT-matrix-based (101 positions) models for nucleotide and amino acid trees, respectively, and bootstrap values were determined using 1000 replicates. Bayesian analysis was conducted to provide further support for the nucleic acid phylogeny using Mr. Bayes (Ronquist et al., 2011), using a GTR substitution model with gamma distributed rate variation, and branches supported by this analysis are indicated with black squares (Fig. 1A). Maximum-likelihood trees of 18S rRNA genes were constructed using PhyML (Guindon et al. 2010) to analyze phylogeny of the UCYN-A hosts and other publically-available related sequences. Phylogenetic trees were visualized in Tree View (Page 1996).

Acknowledgements

We thank Mark Hildebrand of the Scripps Institution of Oceanography for use of the Influx flow cytometer during our sampling sessions (NSF MRI 092368). Tracy Serros, Ellen Campbell, and Ariel Rabines helped with sample processing. This work was funded by a Gordon and Betty Moore Foundation Marine Investigator award (JPZ) and the National Science Foundation Science Center for Microbial Oceanography Research and Education (C-MORE, grant #EF-0424599). Methods were developed through the MEGAMER facility at UC Santa Cruz, which is funded through the Gordon and Betty Moore Foundation. FA and FM were supported by the

Gordon and Betty Moore Foundation MML. AWT is currently an employee of BD Biosciences, the maker of the Influx cell sorter used in this study.

Figure legends

Figure 1. Maximum likelihood trees of representative partial nitrogenase (*nifH)* UCYN-A sequences. Representative sequences were determined by clustering all UCYN-A sequences deposited in NCBI's Genbank database at 97% nucleotide similarity using CD-HIT-EST. A) Nucleic acid tree showing at least three phylogenetically-distinct clusters of UCYN-A. Clades are designated UCYN-A1 (initially characterized at Station ALOHA and includes the UCYN-A genome - (Zehr et al. 2008)), UCYN-A2, and UCYN-A3, and are color-coded. Black text indicates sequences that do not clearly cluster with any of the clades defined in this study. Genbank accession numbers are in parenthesis and the number of clones represented by each sequence is in bold brackets. Bootstrap values are indicated on branches with bootstrap support of >500 out of 1000 permutations. Branches supported by Bayesian analysis are noted with a black triangle. B) Amino acid tree that shows no obvious clustering of UCYN-A sequences. Sequences are color-coded to reflect their clustering in (A).

Figure 2. Distribution and counts of UCYN-A *nifH* sequences from GenBank. A) UCYN-A1, B) UCYN-A2, C) UCYN-A3, and D) other UCYN-A sequences not included in clades defined here. The size of the circle is proportional to the number of sequences that fall into each clade according to the legend in panel D.

448 Figure 3. Comparison of *nifH* gene expression between UCYN-A1 at ALOHA (dotted line)
449 (Church et al. 2005) and UCYN-A2 at SIO (solid line) (This study). Gene expression is
450 calculated relative to gene copy number from DNA (cDNA/DNA). Gene expression at each
451 time-point is normalized to the median gene expression over the entire time-course for each
452 study. Black bars represent dark periods over the diel cycles.
453
454 Figure 4. Flow cytometry analysis of phytoplankton populations at the SIO Pier. (A) Populations
455 present in untreated seawater. UCYN-A2 was detected by qPCR in sort gates A and B and absent
456 from sort gates C, D, E, and F. (B) Populations present in agitated and concentrated seawater
457 (See Methods). UCYN-A2 was detected in the population indicated using qPCR. Standard beads
458 are present for FSC reference only as different data collection settings were utilized for each
459 panel. 3 μm diameter beads in panel A are full-spectrum ultra-bright beads and thus are more
460 fluorescent than the 3 μm diameter YG beads used in panel B. 10 μm diameter beads were the
461 same in both samples (See Experimental Procedures).
462
463 Figure 5. Maximum-likelihood tree showing the phylogeny of the UCYN-A2 host 18S rRNA
464 gene sequences (green box) from single cells relative to other publically-available sequences
465 from isolates (black), environmental sequences (green and blue type), and the UCYN-A1 host
466 sequence (blue box). Sequences amplified from the oligotrophic environments of the North
467 Pacific (Stn. ALOHA) and Eastern South Pacific are typed in blue. Sequences amplified from
468 coastal environments of Japan (*Braarudosphaera* sp.), the Cariaco Basin, and coastal Southern
469 California (SIO Pier) are typed in green.
470

Figure 6. (A) Abundance of UCYN-A2 *nifH* and *Braarudosphaera* 18S rRNA genes at the SIO Pier over three diel cycles. Black bars represent dark periods over the diel cycles.

References

Beaufort, L., I. Probert, T. de Garidel-Thoron, E. M. Bendif, D. Ruiz-Pino, N. Metzl, et al. (2011). "Sensitivity of coccolithophores to carbonate chemistry and ocean acidification." Nature 476(7358): 80-83.

Bench, S. R., P. Heller, I. Frank, M. Arciniega, I. N. Shilova and J. P. Zehr (2013). "Whole genome comparison of six *Crocosphaera watsonii* strains with differing phenotypes." Journal of Phycology.

Bombar, D., P. Heller, P. Sanchez-Baracaldo, B. J. Carter and J. P. Zehr (*Submitted*). "Intriguing genome sequence divergence between two closely related strains of *Candidatus* Atelocyanobacterium thalassa (UCYN-A)." ISME J.

Borsetti, A. M. and F. Cati (1972). "Il nannoplancton calcareo vivente nel Tirreno Centro-Meridionale." Giornale di Geologica 38(395-452).

Carpenter, E. J. and R. A. Foster (2003). Marine Cyanobacterial Symbioses. Cyanobacteria in Symbiosis. A. Rai, B. Bergman and U. Rasmussen, Springer Netherlands**:** 11-17.

Church, M. J., B. D. Jenkins, D. M. Karl and J. P. Zehr (2005). "Vertical distributions of nitrogen-fixing phylotypes at Stn Aloha in the oligotrophic North Pacific Ocean." Aquatic Microbial Ecology 38(1): 3-14.

Church, M. J., C. M. Short, B. D. Jenkins, D. M. Karl and J. P. Zehr (2005). "Temporal Patterns of Nitrogenase Gene (*nifH*) Expression in the Oligotrophic North Pacific Ocean." Appl. Environ. Microbiol. 71(9): 5362-5370.

Coleman, M. L. and S. W. Chisholm (2010). "Ecosystem-specific selection pressures revealed through comparative population genomics." Proceedings of the National Academy of Sciences 107(43): 18634-18639.

Doney, S. C., V. J. Fabry, R. A. Feely and J. A. Kleypas (2009). "Ocean Acidification: The Other CO2 Problem." Annual Review of Marine Science 1(1): 169-192.

Foster, R. A., E. J. Carpenter and B. Bergman (2006). "Unicellular cyanobionts in open ocean dinoflagellates, radiolarians, and tintinnids: Ultrastructural characterization and immuno-localization of phycoerythrin and nitrogenase." Journal of Phycology 42(2): 453-463.

Foster, R. A., M. M. M. Kuypers, T. Vagner, R. W. Paerl, N. Musat and J. P. Zehr (2011). "Nitrogen fixation and transfer in open ocean diatom-cyanobacterial symbioses." ISME J 5(9): 1484-1493.

Frada, M. J., K. D. Bidle, I. Probert and C. de Vargas (2012). "In situ survey of life cycle phases of the coccolithophore *Emiliania huxleyi* (Haptophyta)." Environmental Microbiology 14(6): 1558-1569.

516 Gaarder, K. R. (1954). "Coccolithineae, Silicoflagellatae, Pterospermataceae and other forms from the
517 "Michael Sars" North Atlantic Deep-Sea Expedition 1910." Report on the Scientific Results of the
518 "Michael Sars" North Atlantic Deep-Sea Expedition 2(4): 1-20.

519

520 Goebel, N. L., K. A. Turk, K. M. Achilles, R. Paerl, I. Hewson, A. E. Morrison, et al. (2010). "Abundance
521 and distribution of major groups of diazotrophic cyanobacteria and their potential contribution to N_2
522 fixation in the tropical Atlantic Ocean." Environmental Microbiology 12(12): 3272-3289.

523

524 Guindon, S., J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk and O. Gascuel (2010). "New
525 Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of
526 PhyML 3.0." Systematic Biology 59(3): 307-321.

527

528 Hagino, K., R. Onuma, M. Kawachi and T. Horiguchi (2013). "Discovery of an Endosymbiotic Nitrogen-
529 Fixing Cyanobacterium UCYN-A in *Braarudosphaera bigelowii* (Prymnesiophyceae)." PLoS ONE
530 8(12): e81749.

531

532 Hagino, K., Y. Takano and T. Horiguchi (2009). "Pseudo-cryptic speciation in *Braarudosphaera*
533 *bigelowii* (Gran and Braarud) Deflandre." Marine Micropaleontology 72(3-4): 210-221.

534

535 Halm, H., P. Lam, T. G. Ferdelman, G. Lavik, T. Dittmar, J. LaRoche, et al. (2012). "Heterotrophic
536 organisms dominate nitrogen fixation in the South Pacific Gyre." ISME J 6(6): 1238-1249.

537

Hewson, I., P. H. Moisander, A. E. Morrison and J. P. Zehr (2007). "Diazotrophic bacterioplankton in a coral reef lagoon: phylogeny, diel nitrogenase expression and response to phosphate enrichment." ISME J 1(1): 78-91.

Hilton, J. A., R. A. Foster, H. J. Tripp, B. J. Carter, J. P. Zehr and T. A. Villareal (2013). "Genomic deletions disrupt nitrogen metabolism pathways of a cyanobacterial diatom symbiont." Submitted.

Houdan, A., I. Probert, K. V. Lenning and S. b. Lefebvre (2005). "Comparison of photosynthetic responses in diploid and haploid life-cycle phases of *Emiliania huxleyi* (Prymnesiophyceae)." Marine Ecology Progress Series 292: 139-146.

Houlton, B. Z., Y.-P. Wang, P. M. Vitousek and C. B. Field (2008). "A unifying framework for dinitrogen fixation in the terrestrial biosphere." Nature 454(7202): 327-330.

Hulburt, E. M. (1962). "Phytoplankton in the southwestern Sargasso Sea and North Equatorial Current." Limnology and Oceanography 7: 307-315.

Jardillier, L., M. V. Zubkov, J. Pearman and D. J. Scanlan (2010). "Significant CO_2 fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean." ISME J 4(9): 1180-1192.

Karl, D. M., M. J. Church, J. E. Dore, R. M. Letelier and C. Mahaffey (2012). "Predictable and efficient carbon sequestration in the North Pacific Ocean supported by symbiotic nitrogen fixation." Proceedings of the National Academy of Sciences 109(6): 1842-1849.

Kettler, G. C., A. C. Martiny, K. Huang, J. Zucker, M. L. Coleman, S. Rodrigue, et al. (2007). "Patterns and Implications of Gene Gain and Loss in the Evolution of *Prochlorococcus*." PLoS Genetics 3(12): e231.

Knappertsbusch, M. (1993). "Geographic distribution of living and Holocene coccolithophores in the Mediterranean Sea." Marine Micropaleontology 21: 219-247.

Konno, S., N. Harada, H. Narita and R. W. Jordan (2007). "Living *Braarudosphaera bigelowii* (Gran & Braarud) Deflandre in the Bering Sea." Journal of Nannoplankton Research 29(2): 78-87.

Krupke, A., N. Musat, J. LaRoche, W. Mohr, B. M. Fuchs, R. I. Amann, et al. (2013). "*In situ* identification and N_2 and C fixation rates of uncultivated cyanobacteria populations." Systematic and Applied Microbiology(0).

Lesser, M. P., C. H. Mazel, M. Y. Gorbunov and P. G. Falkowski (2004). "Discovery of Symbiotic Nitrogen-Fixing Cyanobacteria in Corals." Science 305(5686): 997-1000.

Li, W. and A. Godzik (2006). "Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences." Bioinformatics 22(13): 1658-1659.

López-García, P., H. Philippe, F. Gail and D. Moreira (2003). "Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic Ridge." Proceedings of the National Academy of Sciences 100(2): 697-702.

Malmstrom, R. R., A. Coe, G. C. Kettler, A. C. Martiny, J. Frias-Lopez, E. R. Zinser, et al. (2010). "Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific oceans." ISME J 4(10): 1252-1264.

Man-Aharonovich, D., N. Kress, E. B. Zeev, I. Berman-Frank and O. Béjà (2007). "Molecular ecology of nifH genes and transcripts in the eastern Mediterranean Sea." Environmental Microbiology 9(9): 2354-2363.

Medlin, L., H. J. Elwood, S. Stickel and M. L. Sogin (1988). "The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions." Gene 71(2): 491-499.

Moisander, P. H., R. A. Beinart, I. Hewson, A. E. White, K. S. Johnson, C. A. Carlson, et al. (2010). "Unicellular Cyanobacterial Distributions Broaden the Oceanic N_2 Fixation Domain." Science 327(5972): 1512-1514.

Moisander, P. H., R. A. Beinart, M. Voss and J. P. Zehr (2008). "Diversity and abundance of diazotrophic microorganisms in the South China Sea during intermonsoon." ISME J 2(9): 954-967.

604 Montoya, J. P., C. M. Holl, J. P. Zehr, A. Hansen, T. A. Villareal and D. G. Capone (2004). "High rates of
605 N_2 fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean." Nature 430(7003): 1027-1032.

606

607 Page, R. D. M. (1996). "Tree View: An application to display phylogenetic trees on personal computers."
608 Computer applications in the biosciences : CABIOS 12(4): 357-358.

609

610 Prokopowich, C. D., T. R. Gregory and T. J. Crease (2003). "The correlation between rDNA copy number
611 and genome size in eukaryotes." Genome 46(1): 48-50.

612

613 Schneegurt, M. A., D. M. Sherman and L. A. Sherman (1997). "Growth, physiology, and ultrastructure of
614 a diazotrophic cyanobacterium, *Cyanothece* sp. strain ATCC 51142, in mixotrophic and
615 chemoheterotrophic cultures." Journal of Phycology 33(4): 632-642.

616

617 Sunda, W. G. and D. R. Hardison (2010). "Evolutionary tradeoffs among nutrient acquisition, cell size,
618 and grazing defense in marine phytoplankton promote ecosystem stability." Marine Ecology Progress
619 Series 401: 63-76.

620

621 Sunda, W. G. and S. A. Huntsman (1995). "Iron uptake and growth limitation in oceanic and coastal
622 phytoplankton." Marine Chemistry 50(1–4): 189-206.

623

624 Sunda, W. G. and S. A. Huntsman (1997). "Interrelated influence of iron, light and cell size on marine
625 phytoplankton growth." Nature 390(6658): 389-392.

626

627 Takano, Y., K. Hagino, Y. Tanaka, T. Horiguchi and H. Okada (2006). "Phylogenetic affinities of an
628 enigmatic nannoplankton, *Braarudosphaera bigelowii* based on the SSU rDNA sequences." Marine
629 Micropaleontology 60(2): 145-156.

630

631 Thompson, A. W., R. A. Foster, A. Krupke, B. J. Carter, N. Musat, D. Vaulot, et al. (2012). "Unicellular
632 Cyanobacterium Symbiotic with a Single-Celled Eukaryotic Alga." Science 337(6101): 1546-1550.

633

634 Tripp, H. J., S. R. Bench, K. A. Turk, R. A. Foster, B. A. Desany, F. Niazi, et al. (2010). "Metabolic
635 streamlining in an open-ocean nitrogen-fixing cyanobacterium." Nature 464(7285): 90-94.

636

637 Turk-Kubo, K. A., K. M. Achilles, T. R. Serros, M. Ochiai, J. P. Montoya and J. P. Zehr (2012).
638 "Nitrogenase (*nifH*) gene expression in diazotrophic cyanobacteria in the Tropical North Atlantic in
639 response to nutrient amendments." Frontiers in Microbiology 3.

640

641 Turk-Kubo, K. A., M. Karamchandani, D. G. Capone and J. P. Zehr (2013). "The paradox of marine
642 heterotrophic nitrogen fixation: Abundances of heterotrophic diazotrophs do not account for nitrogen
643 fixation rates in the Eastern Tropical South Pacific." Environmental Microbiology: n/a-n/a.

644

645 Turk, K. A., A. P. Rees, J. P. Zehr, N. Pereira, P. Swift, R. Shelley, et al. (2011). "Nitrogen fixation and
646 nitrogenase (nifH) expression in tropical waters of the eastern North Atlantic." ISME J 5(7): 1201-1212.

647

von Dassow, P., H. Ogata, I. Probert, P. Wincker, C. Da Silva, S. Audic, et al. (2009). "Transcriptome analysis of functional differentiation between haploid and diploid cells of *Emiliania huxleyi*, a globally significant photosynthetic calcifying cell." Genome Biology 10(10): 1-33.

Webb, E. A., I. M. Ehrenreich, S. L. Brown, F. W. Valois and J. B. Waterbury (2009). "Phenotypic and genotypic characterization of multiple strains of the diazotrophic cyanobacterium, *Crocosphaera watsonii*, isolated from the open ocean." Environmental Microbiology 11(2): 338-348.

Zani, S., M. T. Mellon, J. L. Collier and J. P. Zehr (2000). "Expression of nifH Genes in Natural Microbial Assemblages in Lake George, New York, Detected by Reverse Transcriptase PCR." Applied and Environmental Microbiology 66(7): 3119-3124.

Zehr, J. P., S. R. Bench, B. J. Carter, I. Hewson, F. Niazi, T. Shi, et al. (2008). "Globally Distributed Uncultivated Oceanic N_2-Fixing Cyanobacteria Lack Oxygenic Photosystem II." Science 322(5904): 1110-1112.

Zehr, J. P. and R. M. Kudela (2011). "Nitrogen cycle of the open ocean: from genes to ecosystems." Ann Rev Mar Sci 3: 197-225.

Zehr, J. P. and L. A. McReynolds (1989). "Use of degenerate oligonucleotides for amplification of the nifH gene from the marine cyanobacterium *Trichodesmium thiebautii*." Applied and Environmental Microbiology 55(10): 2522-2526.

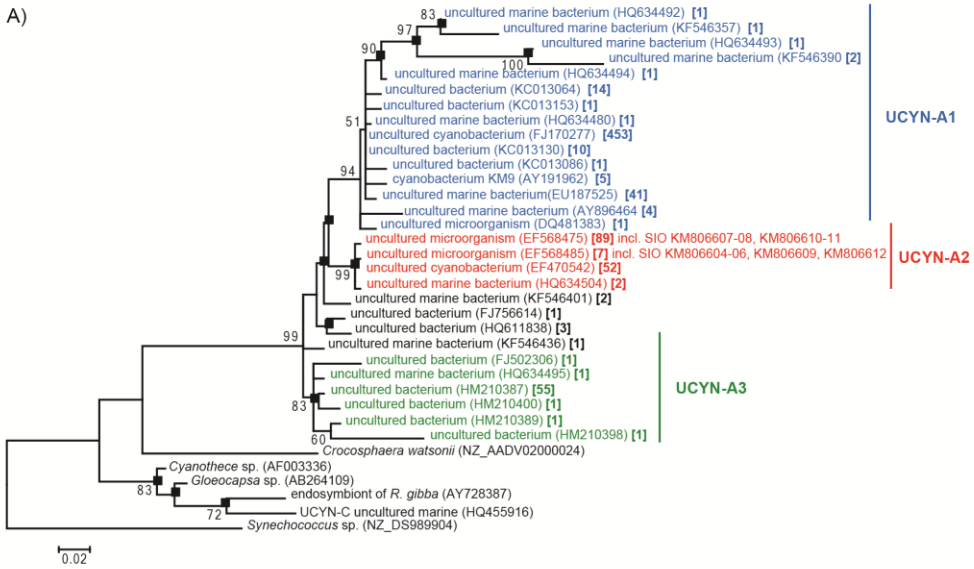
671 Zhu, F., R. Massana, F. Not, D. Marie and D. Vaulot (2005). "Mapping of picoeucaryotes in marine
672 ecosystems with quantitative PCR of the 18S rRNA gene." FEMS Microbiology Ecology 52(1): 79-92.

673

674

675

Figure 1


A)
uncultured marine bacterium (HQ634492) [1]
uncultured marine bacterium (KF546357) [1]
uncultured marine bacterium (HQ634493) [1]
uncultured marine bacterium (KF546390) [2]
uncultured marine bacterium (HQ634494) [1]
uncultured bacterium (KC013064) [14]
uncultured bacterium (KC013153) [1]
uncultured marine bacterium (HQ634480) [1]
uncultured cyanobacterium (FJ170277) [453]
uncultured bacterium (KC013130) [10]
uncultured bacterium (KC013086) [1]
cyanobacterium KM9 (AY191962) [5]
uncultured marine bacterium(EU187525) [41]
uncultured marine bacterium (AY896464 [4]
uncultured microorganism (DQ481383) [1]
uncultured microorganism (EF568475) [89] incl. SIO KM806607-08, KM806610-11
uncultured microorganism (EF568485) [7] incl. SIO KM806604-06, KM806609, KM806612
uncultured cyanobacterium (EF470542) [52]
uncultured marine bacterium (HQ634504) [2]
uncultured marine bacterium (KF546401) [2]
uncultured bacterium (FJ756614) [1]
uncultured bacterium (HQ611838) [3]
uncultured marine bacterium (KF546436) [1]
uncultured bacterium (FJ502306) [1]
uncultured marine bacterium (HQ634495) [1]
uncultured bacterium (HM210387) [55]
uncultured bacterium (HM210400) [1]
uncultured bacterium (HM210389) [1]
uncultured bacterium (HM210398) [1]
Crocosphaera watsonii (NZ_AADV02000024)
Cyanothece sp. (AF003336)
Gloeocapsa sp. (AB264109)
endosymbiont of R. gibba (AY728387)
UCYN-C uncultured marine (HQ455916)
Synechococcus sp. (NZ_DS989904)
UCYN-A1
UCYN-A2
UCYN-A3
0.02

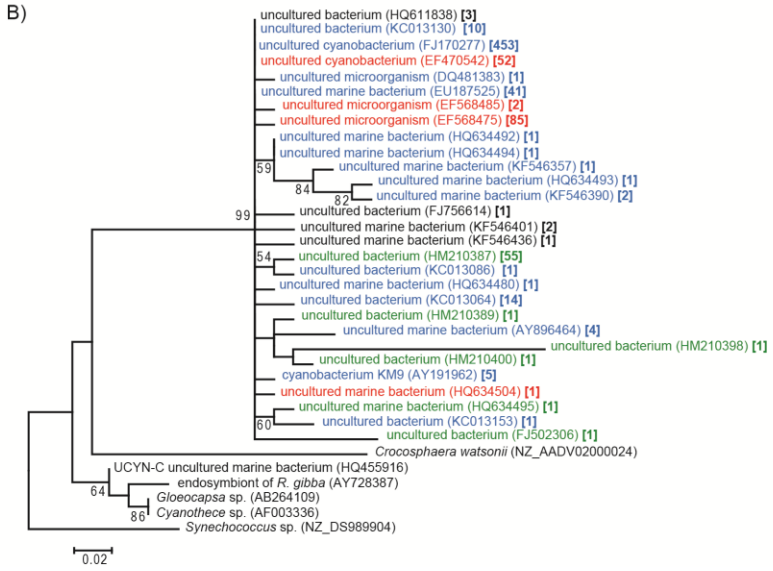

B)
uncultured bacterium (HQ611838) [3]
uncultured bacterium (KC013130) [10]
uncultured cyanobacterium (FJ170277) [453]
uncultured cyanobacterium (EF470542) [52]
uncultured microorganism (DQ481383) [1]
uncultured marine bacterium (EU187525) [41]
uncultured microorganism (EF568485) [2]
uncultured microorganism (EF568475) [85]
uncultured marine bacterium (HQ634492) [1]
uncultured marine bacterium (HQ634494) [1]
uncultured marine bacterium (KF546357) [1]
uncultured marine bacterium (HQ634493) [1]
uncultured marine bacterium (KF546390) [2]
uncultured bacterium (FJ756614) [1]
uncultured marine bacterium (KF546401) [2]
uncultured marine bacterium (KF546436) [1]
uncultured bacterium (HM210387) [55]
uncultured bacterium (KC013086) [1]
uncultured marine bacterium (HQ634480) [1]
uncultured bacterium (KC013064) [14]
uncultured bacterium (HM210389) [1]
uncultured marine bacterium (AY896464) [4]
uncultured bacterium (HM210398) [1]
uncultured bacterium (HM210400) [1]
cyanobacterium KM9 (AY191962) [5]
uncultured marine bacterium (HQ634504) [1]
uncultured marine bacterium (HQ634495) [1]
uncultured bacterium (KC013153) [1]
uncultured bacterium (FJ502306) [1]
Crocosphaera watsonii (NZ_AADV02000024)
UCYN-C uncultured marine bacterium (HQ455916)
endosymbiont of R. gibba (AY728387)
Gloeocapsa sp. (AB264109)
Cyanothece sp. (AF003336)
Synechococcus sp. (NZ_DS989904)
0.02

Figure 2

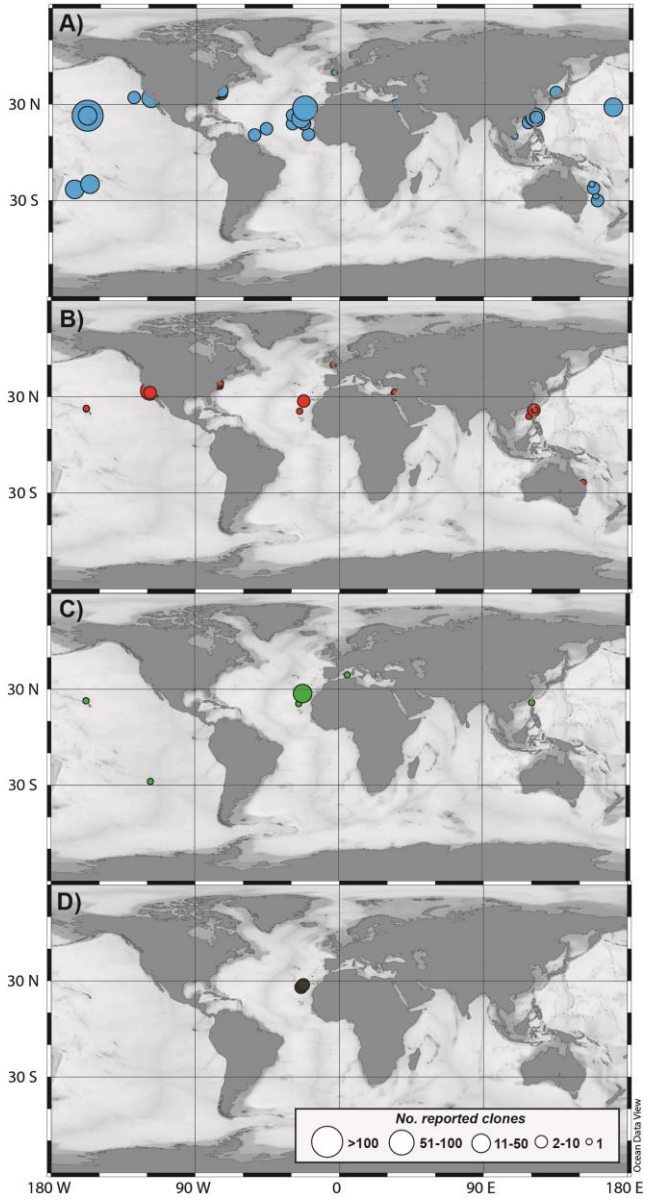

A)
B)
C)
D)
30 N
30 S
180 W
90 W
0
90 E
180 E
No. reported clones
>100
51-100
11-50
2-10
1
Ocean Data View

Figure 3

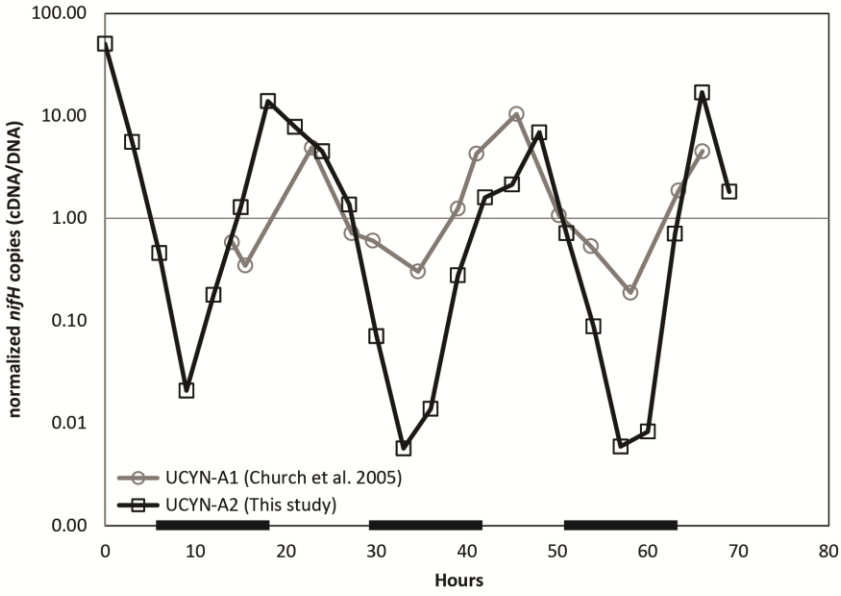

100.00
10.00
1.00
0.10
0.01
0.00
normalized nifH copies (cDNA/DNA)
UCYN-A1 (Church et al. 2005)
UCYN-A2 (This study)
0
10
20
30
40
50
60
70
80
Hours

Figure 4

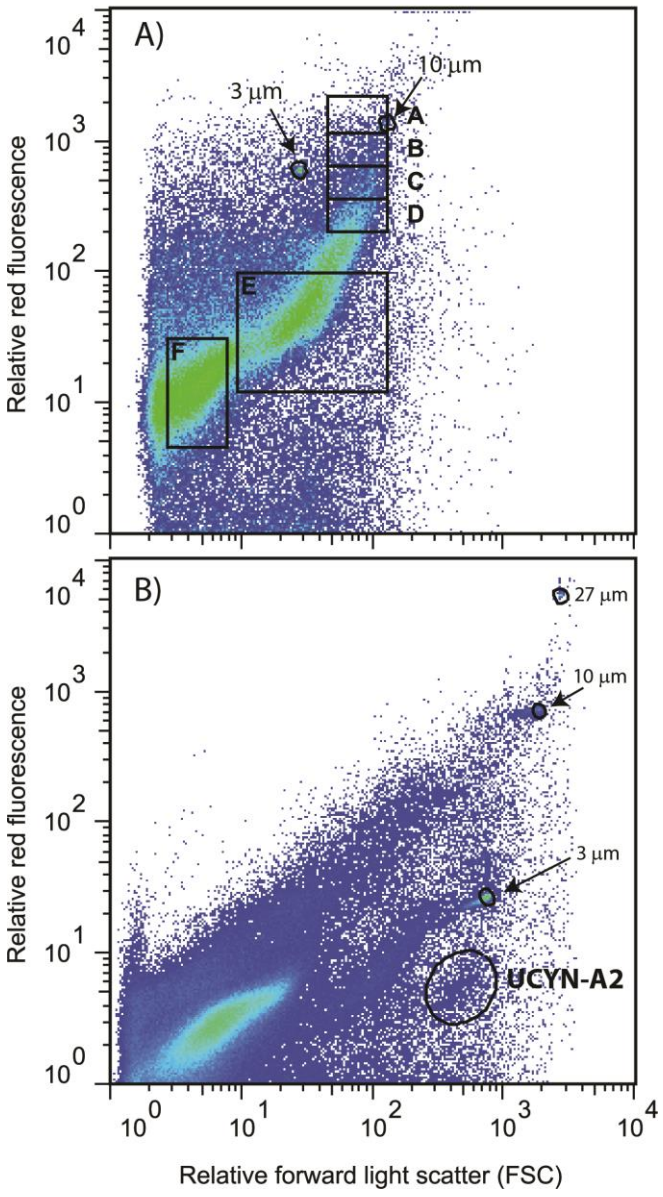

A)
10 µm
3 µm
A
B
C
D
E
F
Relative red fluorescence
B)
27 µm
10 µm
3 µm
UCYN-A2
Relative red fluorescence
Relative forward light scatter (FSC)

Figure 5

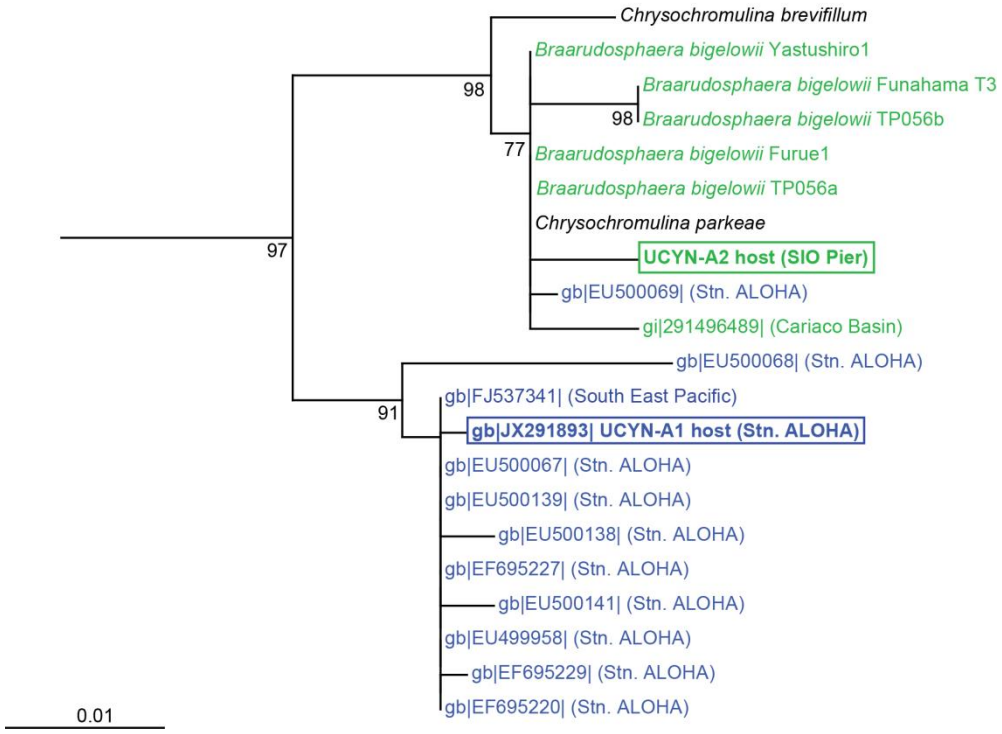
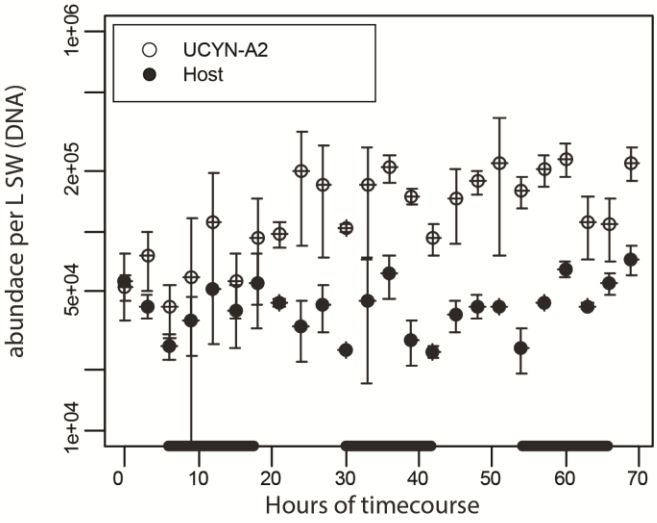

Chrysochromulina brevifillum
Braarudosphaera bigelowii Yastushiro1
Braarudosphaera bigelowii Funahama T3
Braarudosphaera bigelowii TP056b
Braarudosphaera bigelowii Furue1
Braarudosphaera bigelowii TP056a
Chrysochromulina parkeae
UCYN-A2 host (SIO Pier)
gb|EU500069| (Stn. ALOHA)
gi|291496489| (Cariaco Basin)
gb|EU500068| (Stn. ALOHA)
gb|FJ537341| (South East Pacific)
gb|JX291893| UCYN-A1 host (Stn. ALOHA)
gb|EU500067| (Stn. ALOHA)
gb|EU500139| (Stn. ALOHA)
gb|EU500138| (Stn. ALOHA)
gb|EF695227| (Stn. ALOHA)
gb|EU500141| (Stn. ALOHA)
gb|EU499958| (Stn. ALOHA)
gb|EF695229| (Stn. ALOHA)
gb|EF695220| (Stn. ALOHA)
98
98
77
97
91
0.01

Figure 6


UCYN-A2
Host
abundace per L SW (DNA)
1e+04
5e+04
2e+05
1e+06
0
10
20
30
40
50
60
70
Hours of timecourse