Title

Unusual marine unicellular symbiosis with the nitrogen-fixing cyanobacterium UCYN-A

Journal

Nature Microbiology, 2(1)

ISSN

2058-5276

Authors

Zehr, Jonathan P
Shilova, Irina N
Farnelid, Hanna M
et al.

Publication Date

2017

DOI

10.1038/nmicrobiol.2016.214

Peer reviewed

Title: Unusual marine unicellular symbiosis with the nitrogen-fixing cyanobacterium UCYN-A

Jonathan P. Zehr*[1], Irina N. Shilova[1], Hanna M. Farnelid[1,2], Maria del Carmen Muñoz-Marín[1,3], Kendra A. Turk-Kubo[1]

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

[2]Centre for Ecology and Evolution in Microbial Model Systems, Linnaeus University, 392 34 Kalmar, Sweden

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

Abstract

Nitrogen (N_2) fixation, the reduction of N_2 to biologically available nitrogen (N), is an important source of N for terrestrial and aquatic ecosystems. In terrestrial environments, N_2-fixing symbioses involve multicellular plants, but in the marine environment these symbioses occur with unicellular planktonic algae. An unusual symbiosis between an uncultivated unicellular cyanobacterium (UCYN-A) and a haptophyte picoplankton alga was recently discovered in oligotrophic oceans. UCYN-A has a highly reduced genome, and exchanges fixed N for fixed carbon with its host. This symbiosis bears some resemblance to symbioses described in freshwater ecosystems. UCYN-A shares many core genes with the "spheroid bodies" of *Epithemia turgida* and the endosymbionts of the amoeba *Paulinella chromatophora*. UCYN-A is widely distributed, and has diversified into a number of sublineages that could be ecotypes. Many questions remain regarding the physical and genetic mechanisms of the association, but UCYN-A is an intriguing model for contemplating the evolution of N_2-fixing organelles.

Introduction

Humans have used chemical dinitrogen (N_2) fixation, the Haber-Bosch process, for decades to support the nitrogen (N) needs of the human population on Earth (*1*). The great energetic cost of this industrial process makes the process of natural biological N_2 fixation all the more remarkable. Some microorganisms are able to reduce N_2 gas, at substantial costs in ATP and reductant relative to other metabolic enzymatic reactions, using the enzyme nitrogenase (*2*). Nitrogenase, which catalyzes the reduction of N_2 to ammonium, is found in many species of Archaea and Bacteria (including cyanobacteria), but has never been found in a Eukaryote (*2*).

There is only one known way in which N_2 fixation is found in Eukaryotes: in symbiotic interactions between Archaea or Bacteria and Eukaryotic plants or animals (*3, 4*). N_2-fixing symbioses are extremely important as a source of N in both natural and agricultural terrestrial ecosystems, as well as aquatic ecosystems (*1, 5*). This review focuses on a recently discovered

unusual symbiosis between a genetically streamlined marine unicellular cyanobacterium and a single-celled haptophyte algae that is significant in ocean ecology and has implications for understanding the evolution of organelles.

Brief significance of ocean N_2 fixation

N_2 fixation in the oceans was largely ignored for decades, based on Alfred Redfield's arguments that phosphorus was the ultimate limiting nutrient in the oceans (*6*). However, experimental results highlighted that N availability limited productivity in oceans (*7*). More recently it was proposed that there might be an imbalance in inputs and losses of N to the global ocean (*8, 9*), one of the hypotheses being that N_2 fixation had been previously underestimated. Since then, the uncertainties in the N budget have driven intense interest in identifying N_2 fixers and quantifying N_2 fixation in the open oceans.

At the time that N_2 fixation research was re-invigorated, open ocean N_2-fixing microorganisms were believed to be numerically dominated by *Trichodesmium* (*10*) and symbionts of diatoms (*Richelia* that are also filamentous cyanobacteria *11, 12-14*). A number of other N_2-fixing cyanobacteria associated with planktonic Eukaryotes had also been observed microscopically (*15*) but were numerically rare. Molecular biology provided new tools, specifically the polymerase chain reaction (PCR), that did not require cultivation or microscopy to identify N_2-fixing microorganisms in the environment (*16, 17*). Sequencing of PCR-amplified nitrogenase (*nifH*) gene fragments was used to answer the question of whether there were marine N_2-fixing microorganisms other than *Trichodesmium* and the diatom symbionts?

Discovery of unicellular N_2-fixing cyanobacteria group A (UCYN-A)

An unusual cyanobacterium, called UCYN-A, was discovered initially from a short *nifH* gene sequence and, over the ensuing 15 years, was revealed to be widely distributed and to be a very unusual cyanobacterium (Fig. 1). The first *nifH* amplifications from seawater collected in the Atlantic and Pacific Oceans yielded bacterial sequences (which are commonly found in most environments), sequences from *Trichodesmium,* diatom symbionts, and two groups of unicellular cyanobacterial *nifH* sequences (Groups A and B) (*18*) (Fig. 1). This was a surprise, since it was not previously recognized that unicellular N_2-fixing cyanobacteria were significant in the open ocean. *Crocosphaera watsonii,* is a cultivated unicellular cyanobacterium which is genetically identical to the Group B (UCYN-B) *nifH* sequences (*19*). The Group A (UCYN-A) *nifH* sequences were particularly interesting, because they clustered only distantly with the *C. watsonii* sequences and sequences from another unicellular cyanobacterium, *Cyanothece* sp. ATCC 51142 (Fig. 2a). The major surprises lay ahead, as the UCYN-A sequences were not associated with anything that could be visualized in the seawater samples. This was an enigma, since cyanobacteria in that phylogenetic group are typically highly fluorescent, and can usually be identified by epifluorescence microscopy due to the presence of phycoerythrin.

One of the important implications of finding marine unicellular N_2-fixers is that it changed the size-class of organisms that were a source of N in oligotrophic waters. The size of microorganisms fixing N_2 is important since it determines the fate of the fixed N in the food web, and whether the organisms will sink to deep water, sequestering the fixed N (and C fixed by the

host) in the deep ocean (*20*). After the discovery of UCYN-A and UCYN-B, N_2 fixation rates were measured in different size classes to determine whether the newly discovered organisms were significant. Many studies reported substantial, if not most, of the N_2 fixation in the "small" size class dominated by unicellular N_2-fixers (*21-23*).

UCYN-A has been detected in a large number of ocean regions mostly based on cultivation-independent PCR or quantitative PCR (qPCR) approaches targeting the *nifH* gene (*24-33*) (also see these recent reviews *34, 35*), but also fluorescent *in situ* hybridization (FISH) (*36-39*), suggesting UCYN-A is of potential global importance. Initial studies amplified the 16S rRNA gene using generic UCYN (unicellular cyanobacteria, N_2-fixing) PCR primers, and the use of these primer sequences for FISH probes ultimately led to the first real microscopic images of UCYN-A (*25, 36*) (Figs. 1, 3a). However, since the 16S rRNA gene sequence of UCYN-A was not yet known, FISH probes used in these studies cross-hybridized with a number of similar organisms, including UCYN-B (*25, 36*). Since that time, following sequencing of the genome, studies using UCYN-A-specific FISH (*39-43*), metagenomics and metatranscriptomics (*43*) and qPCR approaches have yielded data showing the extensive near global distribution of UCYN-A (Fig. 1) (*35*).

Hints of an unusual physiology in the uncultivated UCYN-A

Despite the inability to cultivate or visualize UCYN-A, experiments at sea focused on developing techniques for quantifying UCYN-A abundances using qPCR, and evaluating the activity by examining the daily pattern of *nifH* gene transcription using quantitative reverse-transcriptase PCR (qRT-PCR) (Fig. 1). The results of qRT-PCR assays for UCYN-A in seawater were enigmatic. Since nitrogenase proteins and enzyme activity are extremely sensitive to oxygen, cyanobacteria separate photosynthesis and N_2-fixing activities spatially (with specialized cells called heterocysts, such as in the diatom symbiont *Richelia*) or temporally by fixing N_2 at night, when photosystem II, the oxygen-evolving component of the photosynthetic apparatus, is not active (*44, 45*). Since UCYN-A and UCYN-B are unicellular, they cannot have heterocysts, and so they should express the N_2-fixing apparatus at night, opposite the pattern of photosynthetic activity which only occurs during the day. Church et al. (*46*) showed that UCYN-B had the expected night-time *nifH* gene expression, but the uncultivated UCYN-A had the highest levels of *nifH* transcripts during the day when it was presumed to be evolving oxygen through photosynthesis. This enigma remained a mystery for a number of years.

Genome sequencing of cyanobacterium UCYN-A resolves enigma

The first glimpse into the genome was obtained from mixed populations in seawater using a flow cytometer to sort cells into a variety of size and fluorescence bins and then identifying the cell populations containing UCYN-A *nifH* by qPCR (Figs. 1, 4) (*47*). The sorted UCYN-A cell populations were used for whole genome amplification and high throughput DNA sequencing (*47*). The resulting sequences clearly showed that the UCYN-A genome was unusual, since it lacked all of the genes for photosystem II (PSII), the oxygen-evolving part of the photosynthetic apparatus. The streamlined genome partially explained the enigma that UCYN-A expressed nitrogenase during the light, since it was now known that UCYN-A did not evolve O_2. More surprises were found when the complete genome was closed (*48*) (Figs. 1, 4). Along with a lack

of PSII, the UCYN-A genome was found to lack Rubisco, the entire tricarboxylic acid (TCA) cycle and a variety of other metabolic pathways. Intriguingly, the UCYN-A genome was shown to be so streamlined as to be missing the characteristics of the group Cyanobacteria, and yet, based on gene sequences, was evolutionarily related to them (Fig. 2a).

The finding of so many missing metabolic pathways indicated that UCYN-A was most likely a symbiont with an organism that was somehow missed in the sample collection and analysis. It was determined that the symbiosis was so fragile that the partners were separated by conventional filtration used to concentrate samples (*40*). Analysis of raw water by flow cytometry showed that the UCYN-A was associated with photosynthetic picoeukaryote cells (Figs. 1, 4), which are diverse, abundant, small species that are only a few micrometers in diameter. A single cell approach was then required to identify the symbiotic partner among the diverse picoeukaryote species. By analyzing single picoeukaryote cells sorted by flow cytometry followed by the 18S rRNA gene amplification and sequencing, the haptophyte partner was identified (Figs. 1, 4) (*40*). The partner cell was closely related to sequences from *Braarudosphaera bigelowii*, a calcifying haptophyte that itself was uncultivated, and most closely related to a sequence amplified from sorted picoplankton cells in a study in the South Pacific gyre (*49*).

With the availability of the genome sequence, a UCYN-A specific 16S rRNA gene FISH probe was designed and the symbiosis was specifically visualized for the first time (*39, 40*) (Fig. 3b, c, d). UCYN-A was given the Candidatus tentative name *Candidatus* Atelocyanobacterium thalassa, meaning incomplete marine cyanobacterium (*40*). Thompson, Foster et al. (*40*) and Krupke et al. (*50*) were able to use stable isotope experiments and nanoSIMS to show that UCYN-A fixed $^{15}N_2$, and that it rapidly exchanged N with the partner haptophyte (Fig. 3b). In return, $H^{13}CO_3^-$ was fixed by the photosynthetic partner and transferred to UCYN-A (*40, 50*). Since the UCYN-A cell was labeled with ^{13}C, and the genome sequence shows carbon (C) fixation pathways are absent in UCYN-A, the isotope experiments and nanoSIMS results clearly demonstrated that the symbiosis was based on the exchange of N and C between the haptophyte and UCYN-A.

Discovery of diversity of UCYN-A and host lineages

Once the UCYN-A symbiosis was discovered, it was not clear whether it was a single partnership between two microorganisms, or there was genetic diversity among UCYN-A strains and perhaps even among the hosts that it partnered with. It was also not known whether there might be UCYN-A relatives that did not have the extreme genome streamlining. There are several reports of possible associations of UCYN-A with other hosts (*41*) or as free-living cells (*36, 41, 51*). About the same time that the UCYN-A symbiosis was discovered, Hagino et al. (*52*) were studying *B. bigelowii* in Japanese coastal waters and observed a small inclusion by transmission electron microscopy (TEM) inside the calcified cell (Fig. 1). PCR amplification of 16S rRNA genes confirmed that UCYN-A was associated with *B. bigelowii* in coastal Japan waters. Also around this time, a second UCYN-A, called UCYN-A2, was detected off of Scripps pier near San Diego, California (*53*) (Fig. 1). UCYN-A2 had the almost all of the same gene deletions in the genome as the original UCYN-A1 genome (*54*), but the DNA and amino acid sequences were surprisingly divergent (average 86% amino acid sequence identity). The

159 genomes of UCYN-A2 and UCYN-A1 also have a number of unique to each strain genes that
160 might contribute to ecologically-relevant physiological differences (*54*). The host of UCYN-A2
161 that was identified by the 18S rRNA gene sequence, was closely related to the UCYN-A1 host,
162 but was even more closely related to the strain from coastal Japanese waters, rather than the
163 sequence from the South Pacific Ocean (*53*). Recently, sequences from UCYN-A and the
164 identified hosts were also found in the large ocean metagenomic and metatranscriptomic datasets
165 from the TARA and MALASPINA oceanographic cruises (*42, 43*) (Fig. 1). Cornejo-Castillo et
166 al. (*43*) confirmed the genome sequence of UCYN-A1 and detected at least one other genome in
167 the South Atlantic Ocean, very similar to the UCYN-A2 genome that was found at Scripps pier
168 in the Pacific Ocean. New FISH probes developed in these studies gave greater resolution to the
169 images of UCYN-A and at least two different hosts (*42, 43*) (Fig. 3g, h). Since multiple strains
170 with genetically distinct differences are now known, a careful analysis of UCYN-A sequences
171 amplified from the environment indicate that there can be multiple subgroups with as yet
172 unknown genomic differences and ecological implications (*35, 40*).

173 The two sublineages that have been substantively characterized thus far, UCYN-A1 and UCYN-
174 A2, are distinctly different in size, but more interestingly, in the number of associated UCYN-A2
175 per host (*42, 43, 53*) (Fig. 3g, h). The UCYN-A2 host is larger (4-10 μm) and appears to have a
176 cluster of UCYN-A cells per haptophyte cell (*43*, *55*). Furthermore, different UCYN-A
177 sublineages have different cell-specific rates of N_2 fixation (*53, 56*).

178 It is now clear that there are distinct pairs of hosts and UCYN-A strains, although there may be
179 yet more unexplored diversity among hosts and UCYN-A, and unknown specificity of
180 associations (Fig. 2b, c). It is not yet known whether characteristics such as size or number of
181 UCYN-A cells per haptophyte partner are reliable defining characteristics for different
182 sublineages. Even more intriguing is whether the different genotypes correspond to different
183 ecotypes that inhabit different habitats or regions of the oceans.

184 **Relationship of UCYN-A symbioses to other symbioses**

185 UCYN-A is the first N_2-fixing symbiont found in the haptophyta (Prymnesiophytes), but there
186 are a number of known symbioses that have some similarities. Known unicellular N_2-fixing
187 symbioses exhibit a spectrum with respect to the closeness of the physical relationship between
188 the partners and associated genome reductions (*53*). The unicellular cyanobacterium
189 *Crocosphaera* has been observed associated with centric diatoms (*57*) (and possibly many other
190 protists), but sequenced genomes of *Crocosphaera* strains are not metabolically streamlined as is
191 UCYN-A (*58*). The heterocyst-forming *Nostoc azollae*, which is an extracellular symbiont of the
192 water fern that inhabits specialized cavities in the plant, has a reduced genome but not as greatly
193 reduced as that of UCYN-A (*59*). In addition, there are other symbioses between cyanobacteria
194 and protists, in particular there are several marine filamentous heterocyst-forming cyanobacteria
195 related to *Richelia* that live within the frustules of diatoms (e.g. *Rhizosolenia* and *Hemiaulus*)
196 (*14, 60, 61*). The most similar symbioses to UCYN-A are the symbioses between freshwater
197 diatoms (of the genera *Rhopalodia* and *Epithemia* and others) and the "spheroid bodies" that are
198 evolutionarily related to cyanobacteria within the same broad phylogenetic group as UCYN-A
199 (*62*). The genome of the spheroid body that lives within the frustule of the diatom *Epithemia*
200 *turgida* was recently sequenced (*63*), and the genome reduction in the *E. turgida* symbiont bears

some similarities to that of UCYN-A, but also has substantive differences, specifically in the gene content (Fig. 5). The symbiont of the diatom *Rhopalodia gibba* appears to be monophyletic with the *E. turgida* symbiont, and has a similar level of genome reduction, although the closed genome has not been published at the time of this writing (*64*). The diatom symbionts are clearly a different lineage from UCYN-A although they are both phylogenetically related to the unicellular cyanobacterium *Cyanothece* and relatives. UCYN-A is one of a group of unicellular N_2-fixing symbionts, but is unique in being a symbiont with a haptophyte alga.

Mechanisms of UCYN-A symbiosis

Symbiotic interactions are facilitated by close physical association, metabolic interdependencies and sometimes gene exchange between partners. The physical association between partners of the UCYN-A symbiosis is as yet unclear because there are only a few low resolution images based on FISH (*39, 56*) or scanning electron microscopy (SEM) (*41*), and the complete genome sequence of the eukaryote partner has not yet been obtained.

Physical association

Symbiotic partners that rely on metabolite exchange have to be structurally connected either by complete enclosure within cell membranes or by having conduits or transporters that enable specific metabolite transfer (Fig. 6). The nature of the physical association of UCYN-A is still unknown, and determining the physical interactions between the symbiotic cells poses technical challenges. However, the different genes determining cell shape and cell wall biogenesis in UCYN-A1 and UCYN-A2 could suggest different associations with their partners (*54*). Images using FISH (*41*) and SEM (*41*) show closely associated cells, suggesting attachment on the outer surface of the haptophyte cell (Fig. 3e, f). In contrast, the *E. turgida* symbiont is an endosymbiont, and the spheroid body is surrounded by membranes, although there is not a continuous connection to the host cytoplasm (*65*) (Fig.6). The spheroid bodies of *Paulinella* (chromatophores) are also intracellular and bounded by membranes (Fig. 6). It will be important to determine whether UCYN-A is surrounded by the host membrane (and thus a true endosymbiont) or is attached to the external surface. If attached on the surface, then there must be molecular mechanisms to maintain attachment and to transfer metabolites (Fig. 6). The mechanism of association must prevent oxygen from inhibiting UCYN-A N_2 fixation, while also exchanging fixed N for fixed C.

Hagino et al. (*52, 66*) were able to visualize *B. bigelowii* in Japan coastal waters because of the presence of unique calcareous plates. TEM images showed 1-2 inclusions of what appears to be UCYN-A (based on PCR amplification of parallel samples (*52*)), within the eukaryote cell *(*Fig. 3i, j, *66*)). These results are puzzling, since they contrast with FISH images that suggest one cell (or several cells) attached to the surface of the eukaryote (*40, 41, 56*). Furthermore, in at least some of the associations, the partners can easily be dislodged from each other by simple filtration (*40*) which would not be true of an intracellular body. The contradictory results might be because the calcified *B. bigelowii* is a distinct life stage (these organisms are known to have complex life cycles) or because the different haptophyte species/strains have different mechanisms of association (and different physical structures).

Metabolic dependency

Close cellular symbiotic interactions can be facilitated by, or require, physical contact and metabolic exchanges (Fig. 6). The specific UCYN-A genome deletions involved in metabolic streamlining may hold clues about the mechanisms involved in maintaining symbiosis. N_2 fixation is a common mechanism that drives many symbiotic interactions between bacteria or cyanobacteria and higher plants, where the plant provides fixed C to the N_2-fixing symbiont (*4*). UCYN-A retains the complete suite of N_2 fixation genes. Clearly, N supply from UCYN-A is the advantage for the haptophyte, which lives in oligotrophic oceans, and UCYN-A requires C from its host (*40, 41, 56*). Metabolic dependencies are likely more severe for the cyanobacterium than the host. UCYN-A does not have a TCA cycle, in contrast to the diatom spheroid bodies that have retained part of the cycle (*63*), although it does have a glycolytic pathway. Retained pathways in UCYN-A suggest metabolisms that are critical for maintenance of viability and transport of nutrients. UCYN-A has lost all ammonium transporters from the genome, but has retained phosphorus transporters, which it must use to obtain P from either the partner or the extracellular environment.

There are interesting differences between the diatom symbionts and UCYN-A. In UCYN-A, photosystem I (PSI) is retained, whereas in the *Epithemia* symbionts both photosystems have been lost. This was interpreted by Nakayama et al. (*63*) to mean that UCYN-A is less far on the evolutionary trajectory to endosymbiosis, but it also could mean that PSI activity is advantageous in the marine environment. Many bacteria in the oceans have proteorhodopsins, or anaerobic anoxygenic photosynthetic pathways (*67*), suggesting that supplementation of energy metabolism may be more important in the marine environment than the freshwater environments where the diatom symbionts are found. UCYN-A is missing the phosphate sensor regulon (*phoBR*), the nitrogen regulating protein PII (*glnB*) and any possible ABC-type nitrate/sulfonate/bicarbonate transporter, relative to the *E. turgida* symbiont. In contrast, UCYN-A has retained the Fe (iron) III transport genes (*afuABC*), which must be important in the Fe-limited oligotrophic oceans. Interestingly, UCYN-A and the *E. turgida* spheroid body have retained the NtcA transcription regulator. NtcA is a N-regulatory transcriptional activator that is required for transcription of *nifH* genes in the absence of ammonium in at least some N_2-fixing cyanobacteria (*68*). The lack of Amt transporters in the *E. turgida* spheroid body and UCYN-A may force the *ntcA* gene to be constitutively up-regulated and be involved in stimulating N_2 fixation. It is also possible that NtcA in UCYN-A responds to the availability of carbon skeletons (69), presumably supplied by the host. The metabolic similarities and differences between the freshwater spheroid bodies and the marine planktonic UCYN-A must be critical for selection in their unique habitats, to have survived such extreme genome reduction.

Symbiosis between two unicellular microorganisms requires more than just energy and metabolite exchange. N_2-fixing symbioses in terrestrial systems between bacteria or cyanobacteria and multicellular land plants involve signaling between the N_2-fixer and the host/partner, which stimulates cell division and allows infection and formation of nodules or nodule-like structures (*4*). The requirements for N_2-fixing symbiosis evolution in multicellular plants are not entirely understood (*4*). Presumably unicellular systems must involve much simpler signal transduction pathways and cellular development or modifications. Coordination of growth and division must be carefully coordinated between two unicellular cells, otherwise one

will outgrow the other. Although the numbers of N_2-fixing symbionts per host may partially depend on the degree of N deficiency, as was observed in *Rhopalodia* (*70 and references therein*), the numbers of cells of both partners have to divide in synchrony so as to pass on the partners to daughter cells following mitotic division.

UCYN-A and Evolution

In addition to being ecologically significant, the UCYN-A symbiosis is an intriguing model of microbial interactions, particularly the unicellular interactions that may be similar to events in the early evolution of organelles (*55*). Non-N_2-fixing interactions involving plastids range across a spectrum from grazing and retention of plastids (kleptoplastids) and loose interactions (*71, 72*), to the very specific interactions of organelles (including chloroplasts) (*73*). Many plastids have evolved as secondary endosymbiosis of a unicellular alga that harbors a chloroplast originally derived from a cyanobacterium. The protist *Paulinella* has incorporated an endosymbiotic cyanobacterium and appears to be an example of an intermediate association that is on an evolutionary trajectory to becoming a plastid (*63*). Some of the characteristics believed to be associated with endosymbiosis are genome reduction and transfer of genes to the host nuclear genome through endosymbiotic gene transfer (EGT) (*74, 75*). The *Paulinella* chromatophore has had extensive genome reduction and transferred at least 30 genes to the host genome (*76, 77*). Most of these EGT-derived genes are related to photosynthesis, including components of the PSI reaction center (*76*).

The UCYN-A genome is very small with short intergenic regions and pseudogenes having the properties of genome reduction of endosymbionts (*54*). Genome comparisons among UCYN-A1, the *E. turgida* spheroid body and the *Paulinella* chromatophore showed that 47% of the UCYN-A1 protein-encoding genes are shared with both of the other endosymbionts (core proteins) and another 44% are shared with at least one of the endosymbiont genomes (Fig. 5).

The diatom spheroid bodies are considered obligate endosymbionts, since they are inseparable from the host cells and are synchronized and passed on to daughter cells during host cell division (*65, 78*). In the case of the *Paulinella* chromatophore, it divides synchronously with the amoeboid host suggesting a certain level of host-endosymbiont integration (*79*). It is still unknown how UCYN-A attaches to the host or if the UCYN-A division is orchestrated by the eukaryote (Fig. 4). However, a few micrographs have been observed that show two UCYN-A1 cells per host right before sunset, which suggests that the cyanobacteria UCYN-A divide prior to the host (*42*).

The distinctions between “endosymbiont” and “organelle” have been debated for a long time. One accepted characteristic used to distinguish between organelles (plastid) and endosymbionts is the presence of a double membrane in organelles and whether proteins are imported (*75, 80*) (Fig. 6). Most organellar proteins are encoded by nuclear genes and translated by host ribosomes, in contrast to endosymbionts where all of the cytosolic proteins are encoded by the endosymbiont genome. It has not yet been shown that the *Paulinella* chromatophore possesses a sophisticated protein import apparatus, and it is debated whether it is an endosymbiont or an organelle. Along with the unicellular cyanobacterial freshwater diatom symbionts, UCYN-A are the only N_2-fixing symbionts that bear some resemblance to a plastid, and may be analogous to

the symbiont of *Paulinella*, except that they fix N_2. It has been questioned why an organelle specialized for N_2 fixation, a "nitroplast", has not evolved. The diatom spheroid bodies and UCYN-A may be analogous to such associations.

The UCYN-A symbiosis has been estimated to have diverged almost 100 million years ago (mya) (*43*). The similarity of deletions among UCYN-A sublineages, but high degree of nucleotide sequence divergence suggests that the genome reduction happened prior to divergence of UCYN-A sublineages and their host partners (*54*). The diatom spheroid bodies have been calculated to have evolved much more recently, only 12 mya (*62*). It is not clear what the initial advantage would have been for the cyanobacterium in either case that would select for the evolution of symbiosis with a phototrophic organism. A similar situation exists with filamentous cyanobacterial symbionts that associate with diatoms and live inside the diatom frustule. Although some of these cyanobacteria also have somewhat reduced genomes (*81*), they are not as drastically reduced as that of UCYN-A and retain complete photosystems (both I and II). Similar to UCYN-A, the filamentous cyanobacterial symbionts have lost ammonium transporter genes from the genome, consistent with an obligate N_2-fixing lifestyle (*81*).

Concluding remarks

The UCYN-A and *Rhopalodia/Epithemia* symbioses are important models for N_2 fixation that contrast with the features and requirements of the better-known symbioses of N_2-fixing Bacteria (including cyanobacteria) with multicellular plants and animals. UCYN-A has yet escaped cultivation, but the new information on metabolic needs informed by genomics may aid in successful culture. Regardless, novel techniques have and will continue to provide information on how this symbiosis works. These symbioses are important to understand since they are analogues for the early stages of the transformations from simple cellular interactions to obligate symbioses and organelles. Since there are no known "nitroplasts," these systems may hold clues as to why such organelles have not evolved, or even provide information that would allow the manipulation of N_2-fixing cyanobacteria with plants. Steps forward require successful cultivation and development of novel methods to visualize the membrane structures and metabolite flow between these small, unicellular partners, as well as sequencing of the partner genomes. It should also be noted that UCYN-A was discovered using a targeted approach (Fig. 4), and there may be other such interactions in nature that provide examples of the spectrum of molecular and physical interactions and stages of evolution from symbiosis to organelle.

Correspondence

All correspondence can be directed to Dr. Jonathan Zehr, zehrj@ucsc.edu.

Acknowledgements

We would like to thank Jonathan Magasin for bioinformatics assistance and the personnel of the Hawaiian Ocean Time Series program. This work supported in part by a grant from the Simons Foundation (SCOPE Award ID 329108, J.P.Z.), and the NSF Center for Microbial Oceanography: Research and Education (C-MORE, DBI #0424599). M.M.M. is supported by a

Marie Curie International Outgoing Fellowship within the 7th European Community Framework Programme. H.F. is supported by the Swedish Research Council VR 637-2013-7502.

Author contributions

J.P.Z. conceptualized and drafted the manuscript. I.N.S. performed genome comparisons and prepared accompanying figure. M.M.M. compiled photomicrographs, and prepared conceptual figures. K.T.K. performed phylogenetic analysis and prepared accompanying figure. J.P.Z., I.N.S., H.M.F., M.M.M. and K.T.K. drafted and edited the manuscript and figures. All authors read and approved the final manuscript.

Competing interest

The authors declare no competing financial interests.

References

1. Fowler, D. *et al*. The global nitrogen cycle in the twenty-first century. *Phil. Trans. R. Soc. B*. **368**, 20130164 (2013).
2. Raymond, J., Siefert, J. L., Staples, C. R. & Blankenship, R. E. The natural history of nitrogen fixation. *Mol. Biol. Evol*. **21**, 541-554 (2004).
3. Kneip, C., Lockhart, P., Voss, C. & Maier, U. G. Nitrogen fixation in eukaryotes-new models for symbiosis. *BMC Evol. Biol*. **7**, 55 (2007).
4. Geurts, R., Xiao, T. T. & Reinhold-Hurek, B. What does it take to evolve a nitrogen-fixing endosymbiosis? *Trends Plant Sci*. **21**, 199-208 (2016).
5. Vitousek, P. M. & Howarth R. W. Nitrogen limitation on land and in the sea: How can it occur? *Biogeochem*. **13**, 87-115 (1991).
6. Redfield, A. C. The biological control of chemical factors in the environment. *Am. Sci*. **46**, 205-222 (1958).
7. Ryther, J. & Dunstan, W. Nitrogen, phosphorus, and eutrophication in the coastal marine environment. *Science* **171**, 1008-1013 (1971).
8. Gruber, N. & Sarmiento, J. L. Global patterns of marine nitrogen fixation and denitrification. *Global Biogeochem. Cy*. **11**, 235-266 (1997).
9. Codispoti, L. A. An oceanic fixed nitrogen sink exceeding 400 Tg N a^{-1} vs the concept of homeostasis in the fixed-nitrogen inventory. *Biogeosciences* **4**, 233-253 (2007).
10. Capone, D., Zehr, J., Paerl, H. & Bergman, B. *Trichodesmium*, a globally significant marine cyanobacterium. *Science* **276**, 1221-1229 (1997).
11. Mague, T. H., Mague, F. C. & Holm-Hansen, O. Physiology and chemical composition of nitrogen fixing phytoplankton in the central North Pacific Ocean. *Mar. Biol*. **41**, 213-227 (1977).
12. Villareal, T. A. Nitrogen-fixation by the cyanobacterial symbiont of the diatom genus *Hemiaulus*. *Mar. Ecol. Prog. Ser*. **76**, 201-204 (1991).
13. Villareal, T. A. Marine nitrogen-fixing diatom-cyanobacteria symbioses. In *Marine Pelagic Cyanobacteria: Trichodesmium and other Diazotrophs,* E. J. Carpenter, D. G.

Capone, J. G. Rueter, Eds. (Kluwer Academic Publishers, The Netherlands, 1992), pp. 163-175.
14. Villareal, T. A. Widespread occurrence of the *Hemiaulus*-cyanobacterial symbiosis in the southwest North-Atlantic Ocean. *Bull. Mar. Sci.* **54**, 1-7 (1994).
15. Foster, R. A., Carpenter, E. J. & Bergman, B. Unicellular cyanobionts in open ocean dinoflagellates, radiolarians, and tintinnids: Ultrastructural characterization and immuno-localization of phycoerythrin and nitrogenase. *J. Phycol.* **42**, 453-463 (2006).
16. Zehr, J. & McReynolds, L. Use of degenerate oligonucleotides for amplification of the *nifH* gene from the marine cyanobacterium *Trichodesmium thiebautii*. *Appl. Environ. Microbiol.* **55**, 2522-2526 (1989).
17. Paerl, H. W., Prufert, L. E. & Ambrose, W. W. Contemporaneous N_2 fixation and oxygenic photosynthesis in the nonheterocystous mat-forming cyanobacterium *Lyngbya aestuarii*. *Appl. Environ. Microbiol.* **57**, 3086-3092 (1991).
18. Zehr, J., Mellon, M. & Zani, S. New nitrogen-fixing microorganisms detected in oligotrophic oceans by amplification of nitrogenase (*nifH*) genes. *Appl. Environ. Microbiol.* **64**, 3444-3450 (1998).
19. Zehr, J. *et al*. Unicellular cyanobacteria fix N_2 in the subtropical North Pacific Ocean. *Nature* **412**, 635-638 (2001).
20. Karl, D. M., Church, M. J., Dore, J. E., Letelier, R. M. & Mahaffey, C. Predictable and efficient carbon sequestration in the North Pacific Ocean supported by symbiotic nitrogen fixation. *Proc. Natl. Acad. Sci. USA.* **109**, 1842-1849 (2012).
21. Dore, J. E., Brum, J. R., Tupas, L. M. & Karl, D. M. Seasonal and interannual variability in sources of nitrogen supporting export in the oligotrophic subtropical North Pacific Ocean. *Limnol. Oceanogr.* **47**, 1595-1607 (2002).
22. Montoya, J. P. *et al*. High rates of N_2 fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean. *Nature* **430**, 1027-1031 (2004).
23. Grabowski, M. N. W., Church, M. J. & Karl, D. M. Nitrogen fixation rates and controls at Stn ALOHA. *Aquat. Microb. Ecol.* **52**, 175-183 (2008).
24. Falcón, L. I., Cipriano, F., Chistoserdov, A. Y. & Carpenter, E. J. Diversity of diazotrophic unicellular cyanobacteria in the tropical North Atlantic Ocean. *Appl. Environ. Microbiol.* **68**, 5760-5764 (2002).
25. Mazard, S. L., Fuller, N. J., Orcutt, K. M., Bridle, O. & Scanlan, D. J. PCR analysis of the distribution of unicellular cyanobacterial diazotrophs in the Arabian Sea. *Appl. Environ. Microbiol.* **70**, 7355-7364 (2004).
26. Church, M. J., Bjorkman, K. M., Karl, D. M., Saito, M. A. & Zehr, J. P. Regional distributions of nitrogen-fixing bacteria in the Pacific Ocean. *Limnol. Oceanogr.* **53**, 63-77 (2008).
27. Langlois, R. J., Hummer, D. & LaRoche, J. Abundances and distributions of the dominant *nifH* phylotypes in the Northern Atlantic Ocean. *Appl. Environ. Microbiol.* **74**, 1922-1931 (2008).
28. Moisander, P. H. *et al*. Unicellular cyanobacterial distributions broaden the oceanic N_2 fixation domain. *Science* **327**, 1512-1514 (2010).
29. Halm, H. *et al*. Heterotrophic organisms dominate nitrogen fixation in the South Pacific Gyre. *ISME J.* **6**, 1238-1249 (2012).
30. Agawin, N. S. R. *et al*. Dominance of unicellular cyanobacteria in the diazotrophic community in the Atlantic Ocean. *Limnol. Oceanogr.* **59**, 623-637 (2014).

31. Bentzon-Tilia, M. *et al*. Significant N_2 fixation by heterotrophs, photoheterotrophs and heterocystous cyanobacteria in two temperate estuaries. *ISME J*. **9**, 273-285 (2015).
32. Messer, L. F., Doubell, M., Jeffries, T. C., Brown, M. V. & Seymour, J. R. Prokaryotic and diazotrophic population dynamics within a large oligotrophic inverse estuary. *Aquat. Microb. Ecol*. **74**, 1-15 (2015).
33. Shiozaki, T., Nagata, T., Ijichi, M. & Furuya, K. Nitrogen fixation and the diazotroph community in the temperate coastal region of the northwestern North Pacific. *Biogeosciences* **12**, 4751-4764 (2015).
34. Luo, Y. W. *et al*. Database of diazotrophs in global ocean: abundance, biomass and nitrogen fixation rates. *Earth Syst. Sci. Data* **4**, 47-73 (2012).
35. Farnelid, H., Turk-Kubo, K., Muñoz-Marín, MC. M. & Zehr, J. P. New insights into the ecology of the globally significant uncultured nitrogen-fixing symbiont UCYN-A. *Aquat. Microb. Ecol*. **77**, 125-138 (2016).
36. Biegala, I. C. & Raimbault, P. High abundance of diazotrophic picocyanobacteria (< 3 μm) in a Southwest Pacific coral lagoon. *Aquat. Microb. Ecol*. **51**, 45-53 (2008).
37. Bonnet, S., Biegala, I. C., Dutrieux, P., Slemons, L. O. & Capone, D. G., Nitrogen fixation in the western equatorial Pacific: Rates, diazotrophic cyanobacterial size class distribution, and biogeochemical significance. *Global Biogeochem. Cy*. **23** GB3012, (2009).
38. Le Moal, M. & Biegala, I. C. Diazotrophic unicellular cyanobacteria in the northwestern Mediterranean Sea: A seasonal cycle. *Limnol. Oceanogr*. **54**, 845-855 (2009).
39. Krupke, A. *et al*. In situ identification and N_2 and C fixation rates of uncultivated cyanobacteria populations. *Syst. Appl. Microbiol*. **36**, 259-271 (2013).
40. Thompson, A. W. *et al*. Unicellular cyanobacterium symbiotic with a single-celled eukaryotic alga. *Science* **337**, 1546-1550 (2012).
41. Krupke, A. *et al*. Distribution of a consortium between unicellular algae and the N_2 fixing cyanobacterium UCYN-A in the North Atlantic Ocean. *Environ. Microbiol*. **16**, 3153-3167 (2014).
42. Cabello, A. M. *et al*. Global distribution and vertical patterns of a prymnesiophyte-cyanobacteria obligate symbiosis. *ISME J*. **10**, 693-706 (2016).
43. Cornejo-Castillo, F. M. *et al*. Cyanobacterial symbionts diverged in the late Cretaceous towards lineage-specific nitrogen fixation factories in single-celled phytoplankton. *Nat. Commun*. **7**, 11071 (2016).
44. Fay, P. Oxygen relations of nitrogen fixation in cyanobacteria. *Microbiol. Mol. Biol. Rev*. **56**, 340-373 (1992).
45. Berman-Frank, I., Lundgren, P. & Falkowski, P. Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. *Res. Microbiol*. **154**, 157-164 (2003).
46. Church, M. J., Short, C. M., Jenkins, B. D., Karl, D. M. & Zehr, J. P. Temporal patterns of nitrogenase gene (*nifH*) expression in the oligotrophic North Pacific Ocean. *Appl. Environ. Microbiol*. **71**, 5362-5370 (2005).
47. Zehr, J. P. *et al*. Globally distributed uncultivated oceanic N_2-fixing cyanobacteria lack oxygenic photosystem II. *Science* **322**, 1110-1112 (2008).
48. Tripp, H. J. *et al*. Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium. *Nature* **464**, 90-94 (2010).

49. Shi, X. L., Marie, D., Jardillier, L., Scanlan, D. J. & Vaulot, D. Groups without cultured representatives dominate eukaryotic picophytoplankton in the oligotrophic South East Pacific Ocean. *PLOS ONE* **4**, (2009).
50. Krupke, A. *et al*. The effect of nutrients on carbon and nitrogen fixation by the UCYN-A–haptophyte symbiosis. *ISME J*. **7**, 1635-1647 (2015).
51. Needham, D. M. & Fuhrman, J. A. Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. *Nat. Microbiol*. **1**, 16005 (2016).
52. Hagino, K., Onuma, R., Kawachi, M. & Horiguchi, T. Discovery of an endosymbiotic nitrogen-fixing cyanobacterium UCYN-A in *Braarudosphaera bigelowii* (Prymnesiophyceae). *PLOS ONE* **8**, e81749 (2013).
53. Thompson, A. *et al*. Genetic diversity of the unicellular nitrogen-fixing cyanobacteria UCYN-A and its prymnesiophyte host. *Environ. Microbiol*. **16**, 3238-3249 (2014).
54. Bombar, D., Heller, P., Sanchez-Baracaldo, P., Carter, B. J. & Zehr, J. P. Comparative genomics reveals surprising divergence of two closely related strains of uncultivated UCYN-A cyanobacteria. *ISME J*. **8**, 2530-2542 (2014).
55. Zehr, J. How single cells work together: Are single-celled symbioses organelle evolution in action? *Science* **349**, 1163-1164 (2015).
56. Martínez-Pérez, C. *et al*. The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle. *Nat. Microbiol*. **1**, 16163 (2016).
57. Carpenter, E. J. & Janson, S. Intracellular cyanobacterial symbionts in the marine diatom *Climacodium frauenfeldianum* (Bacillariophyceae). *J. Phycol*. **36**, 540-544 (2000).
58. Bench, S. R., Heller, P., Frank, I. & Arciniega, M. Whole genome comparison of six *Crocosphaera watsonii* strains with differing phenotypes. *J. Phycol*. **49**, 786-801 (2013).
59. Ran, L. *et al*. Genome erosion in a nitrogen-fixing vertically transmitted endosymbiotic multicellular cyanobacterium. *PLOS ONE* **5**, e11486 (2010).
60. Villareal, T. A. Evaluation of nitrogen fixation in the diatom genus *Rhizosolenia* Ehr. in the absence of its cyanobacterial symbiont *Richelia intracellularis* Schmidt. *J. Plankton Res*. **9**, 965-971 (1987).
61. Villareal, T. A. Laboratory culture and preliminary characterization of the nitrogen-fixing *Rhizosolenia-Richelia* symbiosis. *Mar. Ecol*. **11**, 117-132 (1990).
62. Nakayama, T. *et al*. Spheroid bodies in rhopalodiacean diatoms were derived from a single endosymbiotic cyanobacterium. *J. Plant Res*. **124**, 93-97 (2011).
63. Nakayama, T. *et al*. Complete genome of a nonphotosynthetic cyanobacterium in a diatom reveals recent adaptations to an intracellular lifestyle. *Proc. Natl. Acad. Sci. USA*. **111**, 11407-11412 (2014).
64. Trapp, E. M., Adler, S. & Zauner, S. *Rhopalodia gibba* and its endosymbionts as a model for early steps in a cyanobacterial primary endosymbiosis. *Endocytobiosis Cell Res*. **23**, 21-24 (2012).
65. Prechtl, J., Kneip, C., Lockhart, P., Wenderoth, K. & Maier, U. G. Intracellular spheroid bodies of *Rhopalodia gibba* have nitrogen-fixing apparatus of cyanobacterial origin. *Mol. Biol. Evol*. **21**, 1477-1481 (2004).
66. Hagino, K. *et al*. Extracellular calcification of *Braarudosphaera bigelowii* deduced from electron microscopic observations of cell surface structure and elemental composition of pentaliths. *Mar. Micropaleontol*. **125**, 85-94 (2016).
67. Kolber, Z. S. *et al*. Contribution of aerobic photoheterotrophic bacteria to the carbon cycle in the ocean. *Science* **292**, 2492-2495 (2001).

68. Frías, J. E., Flores, E. & Herrero, A. Requirement of the regulatory protein NtcA for the expression of nitrogen assimilation and heterocyst development genes in the cyanobacterium *Anabaena* sp. PCC7120. *Mol. Microbiol.* **14**, 823-832 (1994).
69. Herrero, A., Muro-Pastor, A. M. & Flores, E. Nitrogen Control in Cyanobacteria. *J. Bacteriol.* **183**, 411-425 (2001).
70. Adler, S., Trapp, E. M., Dede, C., Maier, U. G. & Zauner, S. *Rhopalodia gibba* and its endosymbionts as a model for early steps in a cyanobacterial primary endosymbiosis. In *Endosymbiosis*, 167-179, Springer Vienna, Vienna (2013).
71. Amin, S. A. *et al*. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* **522**, 98-101 (2015).
72. Kazamia, E., Helliwell, K. E., Purton, S. & Smith, A. G. How mutualisms arise in phytoplankton communities: building eco-evolutionary principles for aquatic microbes. *Ecol. Lett.* **19**, 810-822 (2016).
73. Green, B. R. Chloroplast genomes of photosynthetic eukaryotes. *Plant J.* **66**, 34-44 (2011).
74. Dyall, S. D., Brown, M. T. & Johnson, P. J. Ancient invasions: from endosymbionts to organelles. *Science* **304**, 253-257 (2004).
75. Nakayama, T. & Archibald, J. M. Evolving a photosynthetic organelle. *BMC Biol.* **10**, 35 (2012).
76. Nowack, E. C. *et al*. Endosymbiotic gene transfer and transcriptional regulation of transferred genes in *Paulinella chromatophora*. *Mol. Biol. Evol.* **28**, 407-422 (2011).
77. Bodyl, A., Mackiewicz, P. & Gagat, P. Organelle evolution: *Paulinella* breaks a paradigm. *Curr. Biol.* **22**, R304-306 (2012).
78. Miyagishima, S. Y. Mechanism of plastid division: from a bacterium to an organelle. *Plant Physiol.* **155**, 1533-1544 (2011).
79. Melkonian, M. & Mollenhauer, D. Robert Lauterborn (1869-1952) and his *Paulinella* chromatophora. *Protist* **156**, 253-262 (2005).
80. Theissen, U. & Martin, W. The difference between organelles and endosymbionts. *Curr. Biol.* **16**, R1016-1017; author reply R1017-1018 (2006).
81. Hilton, J. A. *et al*. Genomic deletions disrupt nitrogen metabolism pathways of a cyanobacterial diatom symbiont. *Nat. Commun.* **4**, 1767 (2013).
82. Church, M. J., Jenkins, B. D., Karl, D. M. & Zehr, J. P. Vertical distributions of nitrogen-fixing phylotypes at Stn ALOHA in the oligotrophic North Pacific Ocean. *Aquat. Microb. Ecol.* **38**, 3-14 (2005).
83. Needoba, J. A., Foster, R. A., Sakamoto, C. & Zehr, J. P. Nitrogen fixation by unicellular diazotrophic cyanobacteria in the temperate oligotrophic North Pacific Ocean. *Limnol. Oceanogr.* **52**: 1317-1327 (2007).
84. Foster, R. A. *et al*. Influence of the Amazon River plume on distributions of free-living and symbiotic cyanobacteria in the western tropical north Atlantic Ocean. *Limnol. Oceanogr.* **52**: 517-532 (2007).
85. Man-Aharonovich, D., Kress, N., Zeev, E. B., Berman-Frank, I. & Béjà, O. Molecular ecology of *nifH* genes and transcripts in the eastern Mediterranean Sea. *Environ. Microbiol.* **9**: 2354-2363 (2007).
86. Hashimoto, R., Yoshida, T., Kuno, S., Nishikawa, T. & Sako, Y. The first assessment of cyanobacterial and diazotrophic diversities in the Japan Sea. *Fish. Sci.* **78**: 1293-1300 (2012).

87. Brown, S. M. & Jenkins, B. D. Profiling gene expression to distinguish the likely active diazotrophs from a sea of genetic potential in marine sediments. *Environ. Microbiol.* **16**, 3128-3142 (2014).
88 Bothe, H., Tripp, H. J. & Zehr, J.P. Unicellular cyanobacteria with a new mode of life: the lack of photosynthetic oxygen evolution allows nitrogen fixation to proceed. *Arch. Microbiol.* **192**, 783-790 (2010).
89. Heller, P., Tripp, H. J., Turk-Kubo, K. & Zehr, J. P. ARBitrator: a software pipeline for on-demand retrieval of auto-curated *nifH* sequences from GenBank. *Bioinformatics* (2014).
90. Messer, L. F. *et al*. High levels of heterogeneity in diazotroph diversity and activity within a putative hotspot for marine nitrogen fixation. *ISME J*. **10**, 1499-1513 (2016).
91 Turk-Kubo, K. A. *et al*. Diazotroph community succession during the VAHINE mesocosm experiment (New Caledonia lagoon). *Biogeosciences* **12**, 7435–7452 (2015).
92. Pruesse, E., Peplies, J. & Glöckner, F. O. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**, 1823-1829 (2012).
93 Ludwig, W. *et al*. ARB: a software environment for sequence data. *Nucleic Acids Res*. **32**, 1363-1371 (2004).
94. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410 (1990).
95. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726-731 (2016).

612 **Figure Legends**

613 **Figure 1.** Timeline detailing major publications leading from discovery of UCYN-A to genome
614 sequencing and visualization of cells, and the detection in different regions of the world's
615 oceans.

616 **Figure 2.** Phylogenetic trees showing the evolutionary relationship between UCYN-A and other
617 cyanobacteria, as well as the microdiversity of UCYN-A and known UCYN-A hosts. (**a**)
618 Neighbor joining tree based on amino acid sequence from partial *nifH* gene fragments from
619 representative UCYN-A sequences and close cyanobacterial relatives. Adapted from Bothe *et al*.
620 (*88*) (**b**) Maximum likelihood tree of UCYN-A *nifH* partial nucleotide gene sequences adapted
621 from Farnelid *et al*. (*35*). Sequences were acquired from both the National Center for
622 Biotechnology Information (NCBI) Genbank nr/nt database at the last update of a curated *nifH*
623 database (March 2015; (*89*)) and next generation sequencing-based studies (*31, 32, 90, 91*).
624 UCYN-A sublineages are labeled, including a newly described sublineage, UCYN-A4, defined
625 almost exclusively by high throughput sequencing data. (**c**) Neighbor joining tree of partial 18S
626 rRNA gene fragments (644 base pairs) from UCYN-A1 and A2 hosts, and close relatives.
627 Sequences were retrieved from the NCBI Genbank nr/nt database in July 2015, aligned using
628 SINA (*92*), phylogenetic analysis was performed in ARB (*93*). Branch lengths were inferred
629 using the Jukes-Cantor correction. Accession numbers are in parenthesis in branch labels.

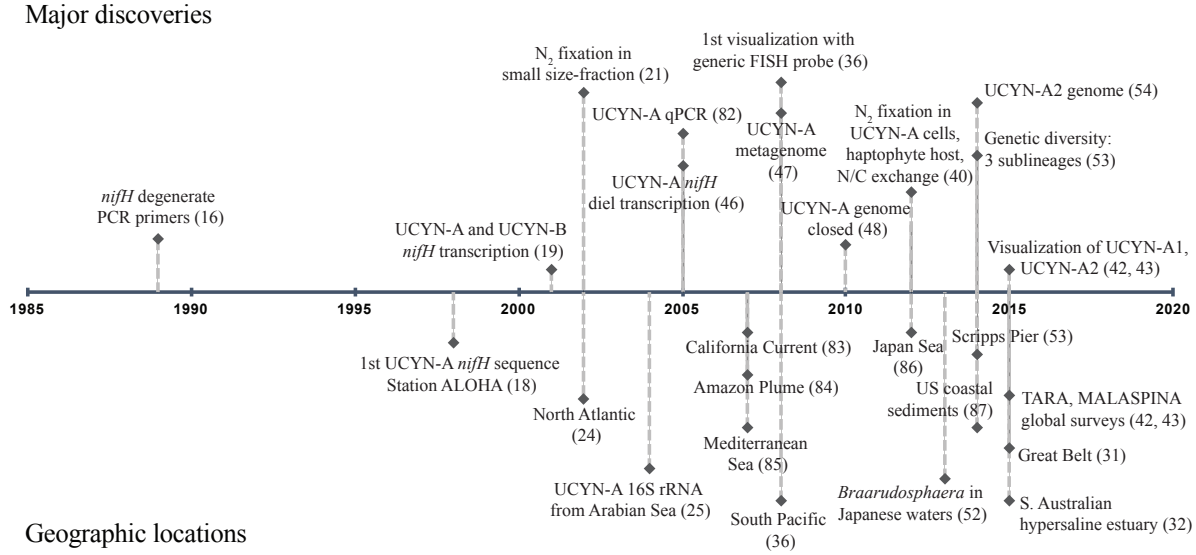
630 **Figure 3.** Visualizations of UCYN-A from key observations between 2008-2016 using a variety
631 of techniques. (**a**) Cells from the 0.2-3 μm size-fraction hybridized with the Nitro821 fluorescent
632 *in situ* hybridization (FISH) probe (generic for all unicellular cyanobacterial N_2-fixers) (*36*). (**b**)
633 UCYN-A cells hybridized with the UCYN-732 probe using catalyzed reporter deposition
634 (CARD)-FISH (green) and DAPI (4´,6-diamidino-2-phenylindole) staining of the partner cell
635 nucleus (blue) (*40*). (**c-d**) UCYN-A cells hybridized with the UCYN-732 probe (green) shows 2
636 different unidentified host cells stained with DAPI (blue), and the chloroplasts of the eukaryotic
637 partner (red) (*39*). (**e-f**) Correlative microscopy of a UCYN-A cell hybridized with the UCYN-
638 A732 probe (green) and associated Haptophyta targeted by the PRYM02 probe (red) in
639 epifluorescence light microscopy (**e**) and electron microscopy (**f**) (*41*). (**g-h**) UCYN-A1
640 identified with the UCYN-A732 probe (red) and prymnesiophyte partner hybridized to the
641 specific probe UPRYM-69 (green) from the MALASPINA Expedition (left panel, **g**) (*42*). *B.*
642 *bigelowii* hybridized to the specific probe UBRADO-69 and UCYN-A2 with the general probe
643 (UCYN-A732) on samples from the TARA Oceans Expedition (right panel, **g**) (*43*). UCYN-A1
644 and its prymnesiophyte partner identified using the specific probes UCYN-A1 732 and UPRYM-
645 69 (upper panel, **h**) and UCYN-A2 targeted by the specific UCYN-A2 753 probe, and the larger
646 prymnesiophyte identified with the UBRADO69 probe (lower panel, **h**) (*35*). (**i-j**) Scanning
647 electron microscopy (SEM) image of *B. bigelowii* (**i**) and transmission electron microscopy
648 (TEM) image of a thin section (**j**) (*66*). (**k-l**) Fluorescence images of UCYN-A1 (**k**) and UCYN-
649 A2 (**l**) with their associated haptophytes after CARD-FISH hybridization with specific probes
650 (*56*). Abbreviations: C - chloroplast; S – spheroid body.

651 **Figure 4**. Methods used to discover the UCYN-A unicellular symbiosis. (**a**) Nitrogenase (*nifH*)
652 genes amplified by PCR from DNA extracted from seawater samples. (**b**, **c**) Screening for
653 UCYN-A in sorted populations of cells by screening for UCYN-A *nifH* by qPCR. (**d**) Genome

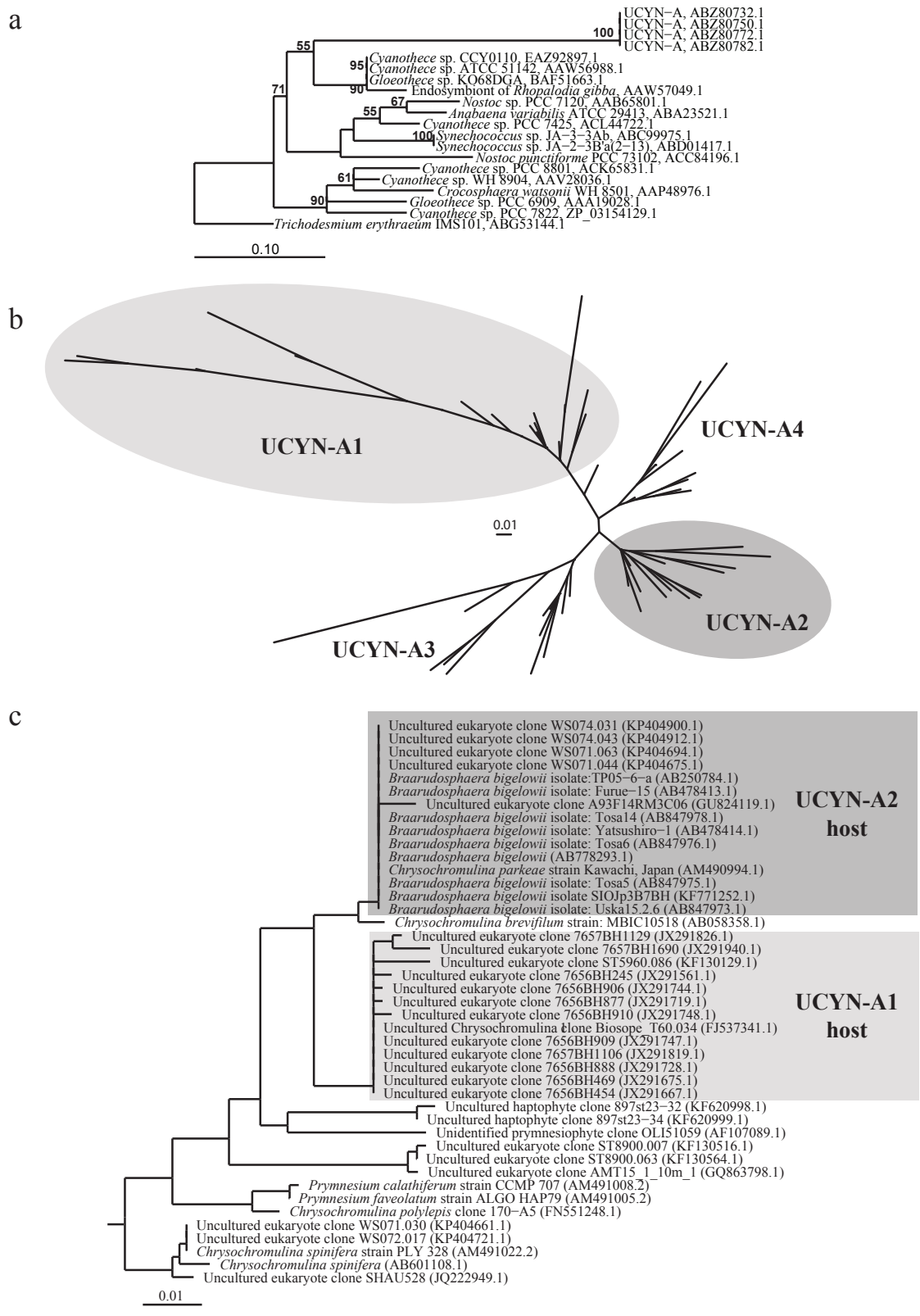
amplification and sequencing from sorted UCYN-A populations. (**e**) Single-cell picophytoplankton cells sorted and screened for UCYN-A *nifH* by qPCR to identify host by 18S rRNA gene PCR amplification. (**f**) Probes designed using the host 18S rRNA gene for FISH/CARD-FISH methods to visualize cells. (**g**) Probes designed from the 16S rRNA gene from genome sequence of UCYN-A for FISH/CARD-FISH methods. (**h**) Cells visualized in seawater samples using FISH/CARD-FISH. Abbreviations: FACS – fluorescence activated cell sorting; *nifH* – nitrogenase gene; PCR – polymerase chain reaction; qPCR – quantitative PCR; FISH – Fluorescence *in situ* hybridization; and CARD-FISH – catalyzed reporter deposition FISH.

Figure 5. Protein comparisons across genomes of UCYN-A1, the *Paulinella* chromatophore and the *E. turgida* spheroid body. (**a**) Venn diagram showing pair-wise shared proteins and the 'core' 533 proteins that were identified as reciprocal match in all pair-wise BLASTP (*94*) searches. The total number of proteins in each organism is noted on the Venn diagram in bold. (**b**) KEGG Orthology (KO) assignment for the core proteins. (**c-e**) Barplots for individual genomes showing KO distribution for proteins that are not in the core set. The genome files for UCYN-A1 and *E. turgida* spheroid body were downloaded from the NCBI Genome database at ftp.ncbi.nlm.nih.gov/genomes/all/ and the *Paulinella* chromatophore genome was downloaded from the NCBI Nucleotide database. Protein fasta files were generated from Genbank files using the BioPython script (gbk_to_faa.py). KO annotation for UCYN-A1 and *E. turgida* spheroid body was downloaded from the KEGG website (http://www.genome.jp/kegg-bin/get_htext) and htext was converted to tabulated files using a custom Perl script. KO for *Paulinella* chromatophore was assigned using BlastKOALA (*95*).

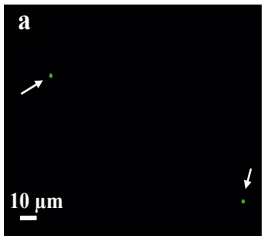
Figure 6. Graphical depictions of plastid evolution and cellular structures in the *Paulinella* and *Epithemia turgida* symbiotic associations and the potential symbiotic interactions in UCYN-A. (**a**) plastid evolution by primary and secondary endosymbiosis. (**b**) *Paulinella chromatophora* and its chromatophore of primary endosymbiotic origin. (**c**) Spheroid bodies integrated into their rhopalodiacean diatom *Epithemia turgida* (**d**) Possible models of symbiotic interactions between UCYN-A and its haptophyte host. Abbreviations: CB - cyanobacterium; N - host nucleus; PL – plastids, and M – mitochondrion; OM – outer membrane; IM – inner membrane.


Major discoveries
N_2 fixation in small size-fraction (21)
1st visualization with generic FISH probe (36)
UCYN-A2 genome (54)
UCYN-A qPCR (82)
UCYN-A metagenome (47)
N_2 fixation in UCYN-A cells, haptophyte host, N/C exchange (40)
Genetic diversity: 3 sublineages (53)
UCYN-A *nifH* diel transcription (46)
nifH degenerate PCR primers (16)
UCYN-A genome closed (48)
UCYN-A and UCYN-B *nifH* transcription (19)
Visualization of UCYN-A1, UCYN-A2 (42, 43)
1985
1990
1995
2000
2005
2010
2015
2020
California Current (83)
Japan Sea (86)
Scripps Pier (53)
1st UCYN-A *nifH* sequence Station ALOHA (18)
Amazon Plume (84)
US coastal sediments (87)
TARA, MALASPINA global surveys (42, 43)
North Atlantic (24)
Mediterranean Sea (85)
Great Belt (31)
UCYN-A 16S rRNA from Arabian Sea (25)
Braarudosphaera in Japanese waters (52)
South Pacific (36)
S. Australian hypersaline estuary (32)
Geographic locations

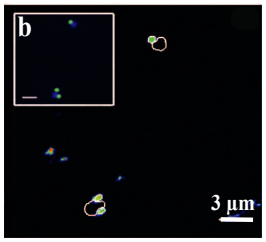
683


a
UCYN−A, ABZ80732.1
UCYN−A, ABZ80750.1
UCYN−A, ABZ80772.1
UCYN−A, ABZ80782.1
100
55
Cyanothece sp. CCY0110, EAZ92897.1
95
Cyanothece sp. ATCC 51142, AAW56988.1
Gloeothece sp. KO68DGA, BAF51663.1
90
71
Endosymbiont of Rhopalodia gibba, AAW57049.1
67
Nostoc sp. PCC 7120, AAB65801.1
55
Anabaena variabilis ATCC 29413, ABA23521.1
Cyanothece sp. PCC 7425, ACL44722.1
100
Synechococcus sp. JA−3−3Ab, ABC99975.1
Synechococcus sp. JA−2−3B'a(2−13), ABD01417.1
Nostoc punctiforme PCC 73102, ACC84196.1
Cyanothece sp. PCC 8801, ACK65831.1
61
Cyanothece sp. WH 8904, AAV28036.1
Crocosphaera watsonii WH 8501, AAP48976.1
90
Gloeothece sp. PCC 6909, AAA19028.1
Cyanothece sp. PCC 7822, ZP_03154129.1
Trichodesmium erythraeum IMS101, ABG53144.1
0.10
b
UCYN-A1
UCYN-A4
0.01
UCYN-A2
UCYN-A3
c
Uncultured eukaryote clone WS074.031 (KP404900.1)
Uncultured eukaryote clone WS074.043 (KP404912.1)
Uncultured eukaryote clone WS071.063 (KP404694.1)
Uncultured eukaryote clone WS071.044 (KP404675.1)
Braarudosphaera bigelowii isolate:TP05−6−a (AB250784.1)
Braarudosphaera bigelowii isolate: Furue−15 (AB478413.1)
Uncultured eukaryote clone A93F14RM3C06 (GU824119.1)
Braarudosphaera bigelowii isolate: Tosa14 (AB847978.1)
Braarudosphaera bigelowii isolate: Yatsushiro−1 (AB478414.1)
Braarudosphaera bigelowii isolate: Tosa6 (AB847976.1)
Braarudosphaera bigelowii (AB778293.1)
Chrysochromulina parkeae strain Kawachi, Japan (AM490994.1)
Braarudosphaera bigelowii isolate: Tosa5 (AB847975.1)
Braarudosphaera bigelowii isolate SIOJp3B7BH (KF771252.1)
Braarudosphaera bigelowii isolate: Uska15.2.6 (AB847973.1)
UCYN-A2 host
Chrysochromulina brevifilum strain: MBIC10518 (AB058358.1)
Uncultured eukaryote clone 7657BH1129 (JX291826.1)
Uncultured eukaryote clone 7657BH1690 (JX291940.1)
Uncultured eukaryote clone ST5960.086 (KF130129.1)
Uncultured eukaryote clone 7656BH245 (JX291561.1)
Uncultured eukaryote clone 7656BH906 (JX291744.1)
Uncultured eukaryote clone 7656BH877 (JX291719.1)
Uncultured eukaryote clone 7656BH910 (JX291748.1)
Uncultured Chrysochromulina clone Biosope_T60.034 (FJ537341.1)
Uncultured eukaryote clone 7656BH909 (JX291747.1)
Uncultured eukaryote clone 7657BH1106 (JX291819.1)
Uncultured eukaryote clone 7656BH888 (JX291728.1)
Uncultured eukaryote clone 7656BH469 (JX291675.1)
Uncultured eukaryote clone 7656BH454 (JX291667.1)
UCYN-A1 host
Uncultured haptophyte clone 897st23−32 (KF620998.1)
Uncultured haptophyte clone 897st23−34 (KF620999.1)
Unidentified prymnesiophyte clone OLI51059 (AF107089.1)
Uncultured eukaryote clone ST8900.007 (KF130516.1)
Uncultured eukaryote clone ST8900.063 (KF130564.1)
Uncultured eukaryote clone AMT15_1_10m_1 (GQ863798.1)
Prymnesium calathiferum strain CCMP 707 (AM491008.2)
Prymnesium faveolatum strain ALGO HAP79 (AM491005.2)
Chrysochromulina polylepis clone 170−A5 (FN551248.1)
Uncultured eukaryote clone WS071.030 (KP404661.1)
Uncultured eukaryote clone WS072.017 (KP404721.1)
Chrysochromulina spinifera strain PLY 328 (AM491022.2)
Chrysochromulina spinifera (AB601108.1)
Uncultured eukaryote clone SHAU528 (JQ222949.1)
0.01

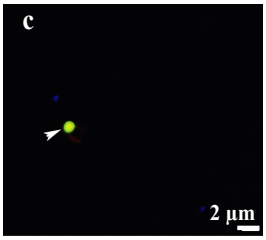
2008: First visualization with generic FISH probe for unicellular diazotrophs (36)

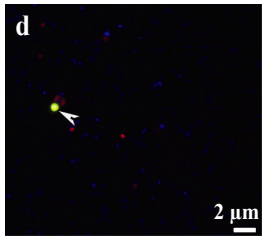


2012: N_2 and CO_2 fixation rates in UCYN-A cells in association with a small Prymnesiophyte at Stn. ALOHA (40)

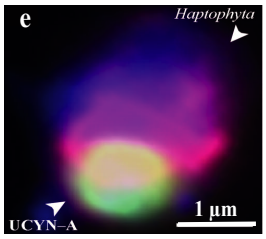


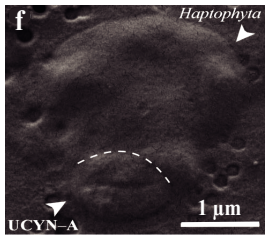
2013: UCYN-A cells attached to two different unidentified hosts in North Atlantic Ocean (39)



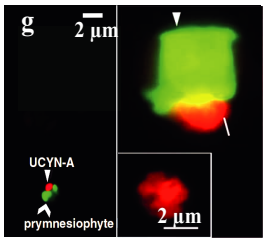


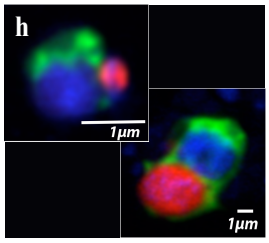
2014: Epifluorescence and SEM showed intimate attachment between both partners (41)



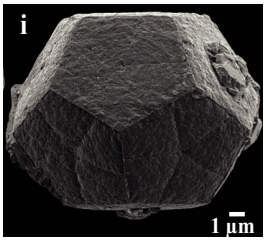


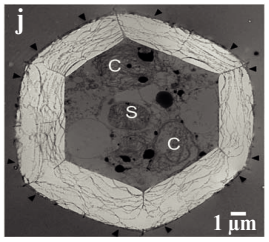
2016: Global distribution of UCYN-A1 and UCYN-A2 in symbiosis with two genetically distinct Prymnesiophytes (35, 42, 43)



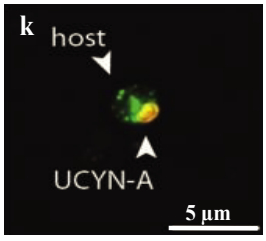


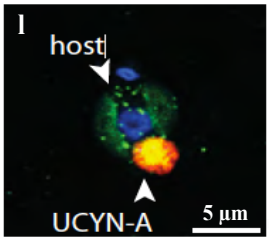
2016: SEM and TEM observations of cell surface structure of *B. bigelowii* in coastal areas (84)

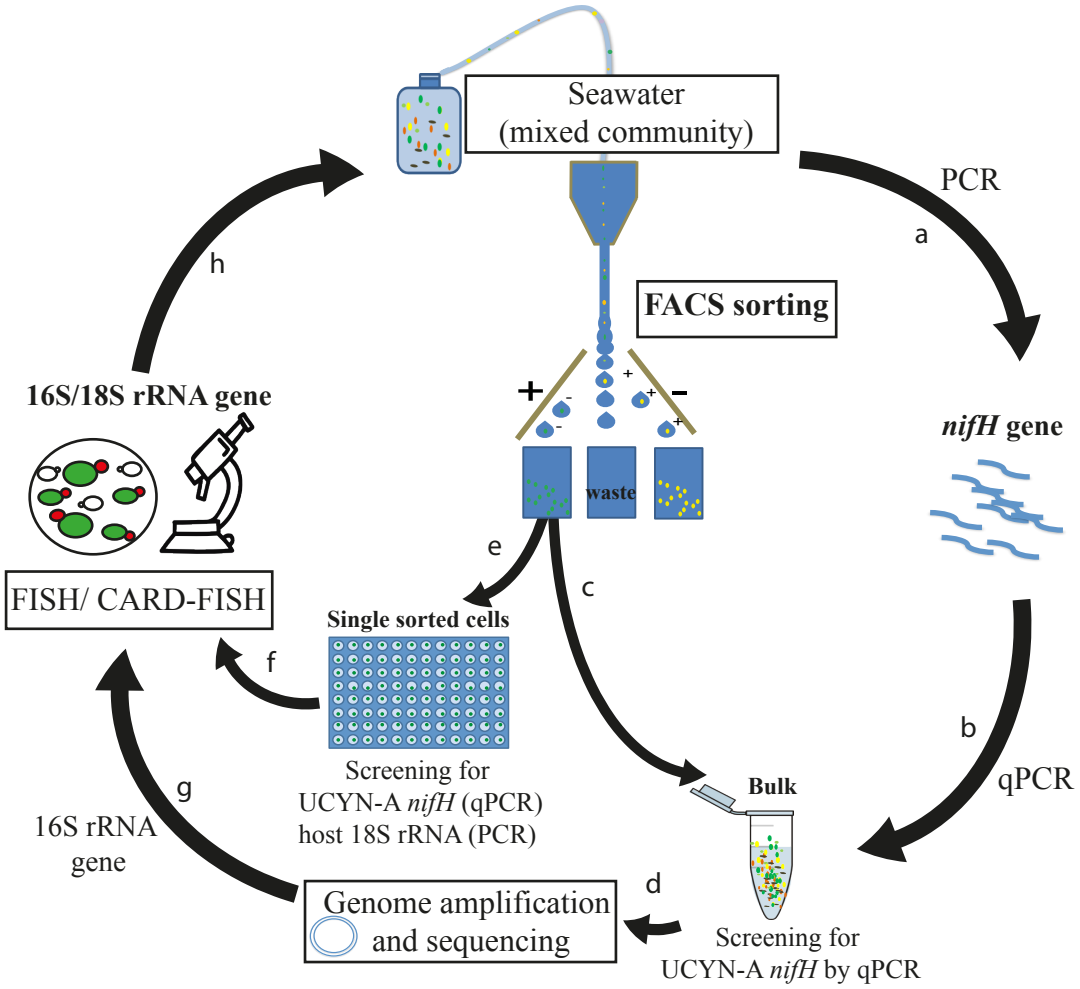




2016: N_2 and CO_2 fixation rates in two different UCYN-A associations in the North Atlantic (56)

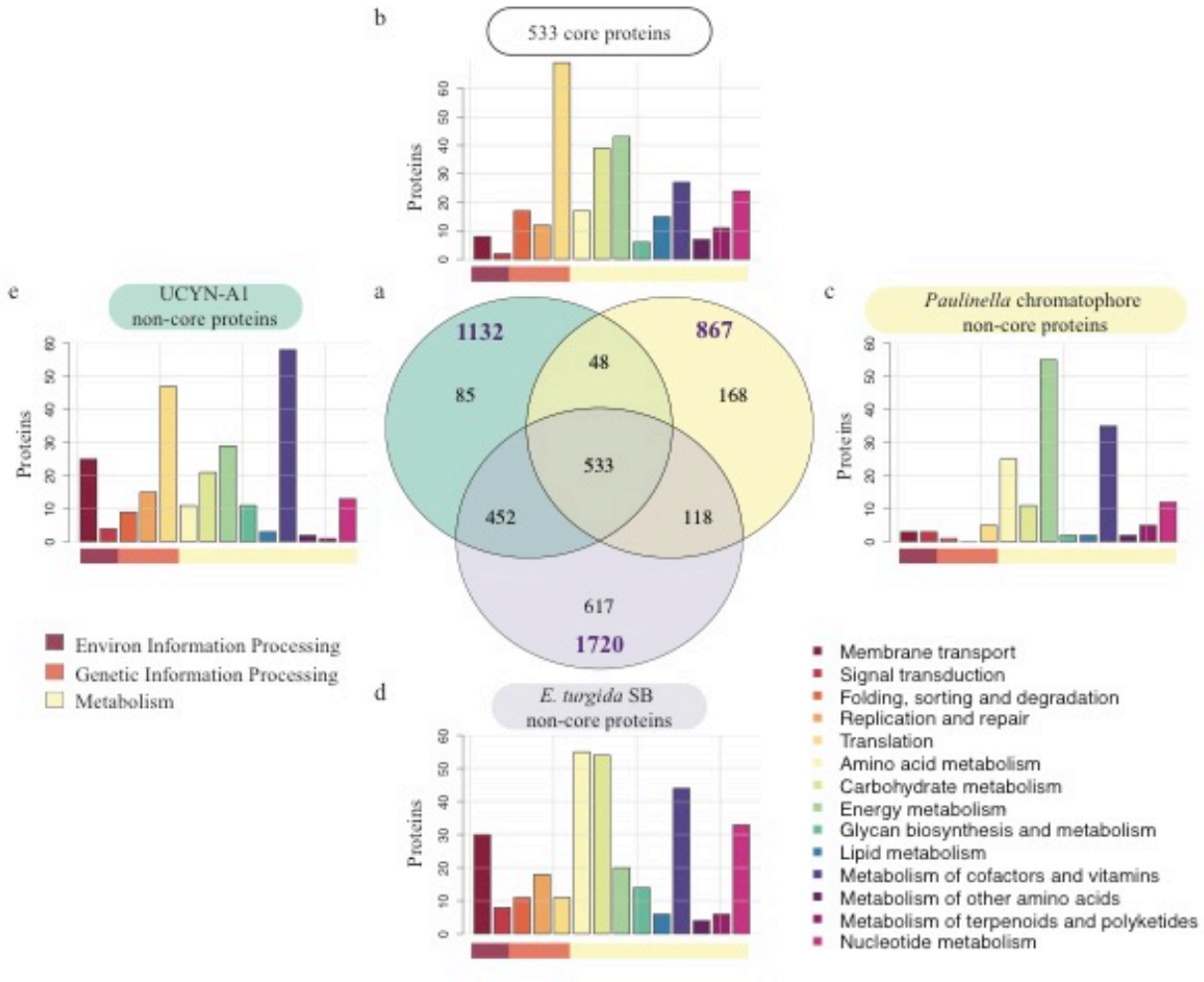




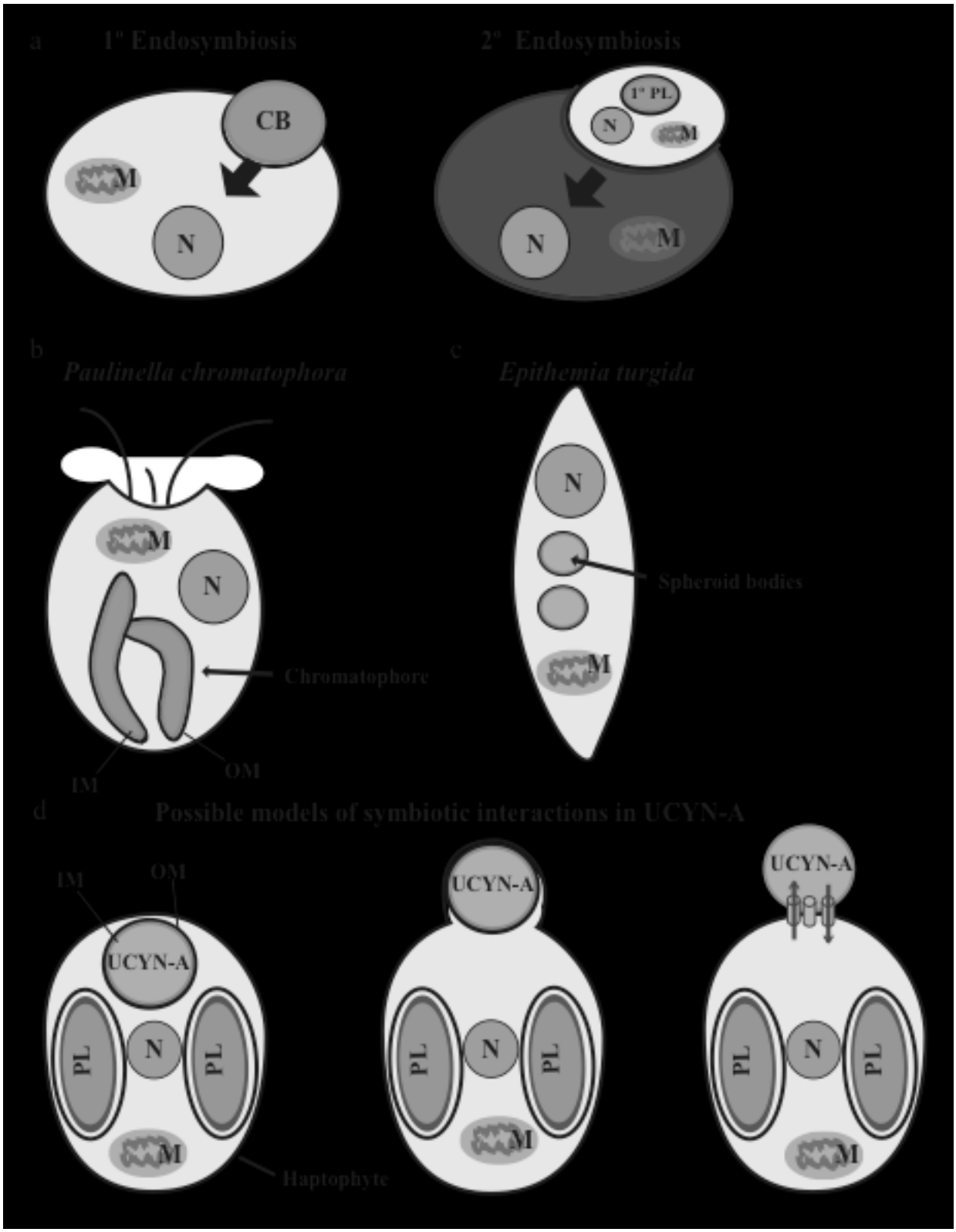

Seawater
(mixed community)
PCR
a
FACS sorting
16S/18S rRNA gene
h
nifH gene
waste
e
c
FISH/ CARD-FISH
Single sorted cells
f
b
qPCR
Bulk
g
Screening for
UCYN-A nifH (qPCR)
host 18S rRNA (PCR)
16S rRNA
gene
d
Genome amplification
and sequencing
Screening for
UCYN-A nifH by qPCR

686

687


b
533 core proteins
Proteins
a
1132
867
48
85
168
533
452
118
617
1720
e
UCYN-A1
non-core proteins
Proteins
c
Paulinella chromatophore
non-core proteins
Proteins
d
E. turgida SB
non-core proteins
Proteins
Environ Information Processing
Genetic Information Processing
Metabolism
Membrane transport
Signal transduction
Folding, sorting and degradation
Replication and repair
Translation
Amino acid metabolism
Carbohydrate metabolism
Energy metabolism
Glycan biosynthesis and metabolism
Lipid metabolism
Metabolism of cofactors and vitamins
Metabolism of other amino acids
Metabolism of terpenoids and polyketides
Nucleotide metabolism

688


a
1° Endosymbiosis
CB
M
N
2° Endosymbiosis
1° PL
N
M
N
M
b
Paulinella chromatophora
M
N
Chromatophore
IM
OM
c
Epithemia turgida
N
Spheroid bodies
M
d
Possible models of symbiotic interactions in UCYN-A
IM
OM
UCYN-A
PL
N
PL
M
Haptophyte
UCYN-A
PL
N
PL
M
UCYN-A
PL
N
PL
M

689