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

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

4

5   [1]Department of Ocean Sciences, University of California, Santa Cruz, CA 95064, USA

6   [2]Centre for Ecology and Evolution in Microbial Model Systems, Linnaeus University, 392 34
7   Kalmar, Sweden

8   [3]Departamento de Bioquímica y Biología Molecular, Edificio Severo Ochoa, Universidad de
9   Córdoba, 14071-Córdoba, Spain

10

11   **Abstract**

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

24   **Introduction**

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

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

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

Brief significance of ocean N$_2$ fixation

40 N$_2$ fixation in the oceans was largely ignored for decades, based on Alfred Redfield's arguments
41 that phosphorus was the ultimate limiting nutrient in the oceans (*6*). However, experimental
42 results highlighted that N availability limited productivity in oceans (*7*). More recently it was
43 proposed that there might be an imbalance in inputs and losses of N to the global ocean (*8, 9*),
44 one of the hypotheses being that N$_2$ fixation had been previously underestimated. Since then, the
45 uncertainties in the N budget have driven intense interest in identifying N$_2$ fixers and quantifying
46 N$_2$ fixation in the open oceans.

47 At the time that N$_2$ fixation research was re-invigorated, open ocean N$_2$-fixing microorganisms
48 were believed to be numerically dominated by *Trichodesmium* (*10*) and symbionts of diatoms
49 (*Richelia* that are also filamentous cyanobacteria *11, 12-14*). A number of other N$_2$-fixing
50 cyanobacteria associated with planktonic Eukaryotes had also been observed microscopically
51 (*15*) but were numerically rare. Molecular biology provided new tools, specifically the
52 polymerase chain reaction (PCR), that did not require cultivation or microscopy to identify N$_2$-
53 fixing microorganisms in the environment (*16, 17*). Sequencing of PCR-amplified nitrogenase
54 (*nifH*) gene fragments was used to answer the question of whether there were marine N$_2$-fixing
55 microorganisms other than *Trichodesmium* and the diatom symbionts?

Discovery of unicellular N$_2$-fixing cyanobacteria group A (UCYN-A)

57 An unusual cyanobacterium, called UCYN-A, was discovered initially from a short *nifH* gene
58 sequence and, over the ensuing 15 years, was revealed to be widely distributed and to be a very
59 unusual cyanobacterium (Fig. 1). The first *nifH* amplifications from seawater collected in the
60 Atlantic and Pacific Oceans yielded bacterial sequences (which are commonly found in most
61 environments), sequences from *Trichodesmium*, diatom symbionts, and two groups of unicellular
62 cyanobacterial *nifH* sequences (Groups A and B) (*18*) (Fig. 1). This was a surprise, since it was
63 not previously recognized that unicellular N$_2$-fixing cyanobacteria were significant in the open
64 ocean. *Crocosphaera watsonii*, is a cultivated unicellular cyanobacterium which is genetically
65 identical to the Group B (UCYN-B) *nifH* sequences (*19*). The Group A (UCYN-A) *nifH*
66 sequences were particularly interesting, because they clustered only distantly with the *C.*
67 *watsonii* sequences and sequences from another unicellular cyanobacterium, *Cyanothece* sp.
68 ATCC 51142 (Fig. 2a). The major surprises lay ahead, as the UCYN-A sequences were not
69 associated with anything that could be visualized in the seawater samples. This was an enigma,
70 since cyanobacteria in that phylogenetic group are typically highly fluorescent, and can usually
71 be identified by epifluorescence microscopy due to the presence of phycoerythrin.

72 One of the important implications of finding marine unicellular N$_2$-fixers is that it changed the
73 size-class of organisms that were a source of N in oligotrophic waters. The size of
74 microorganisms fixing N$_2$ is important since it determines the fate of the fixed N in the food web,
75 and whether the organisms will sink to deep water, sequestering the fixed N (and C fixed by the

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

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

92 **Hints of an unusual physiology in the uncultivated UCYN-A**

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

107 **Genome sequencing of cyanobacterium UCYN-A resolves enigma**

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

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

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

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

Discovery of diversity of UCYN-A and host lineages

146   Once the UCYN-A symbiosis was discovered, it was not clear whether it was a single
147   partnership between two microorganisms, or there was genetic diversity among UCYN-A strains
148   and perhaps even among the hosts that it partnered with. It was also not known whether there
149   might be UCYN-A relatives that did not have the extreme genome streamlining. There are
150   several reports of possible associations of UCYN-A with other hosts (41) or as free-living cells
151   (36, 41, 51). About the same time that the UCYN-A symbiosis was discovered, Hagino et al.
152   (52) were studying *B. bigelowii* in Japanese coastal waters and observed a small inclusion by
153   transmission electron microscopy (TEM) inside the calcified cell (Fig. 1). PCR amplification of
154   16S rRNA genes confirmed that UCYN-A was associated with *B. bigelowii* in coastal Japan
155   waters. Also around this time, a second UCYN-A, called UCYN-A2, was detected off of Scripps
156   pier near San Diego, California (53) (Fig. 1). UCYN-A2 had the almost all of the same gene
157   deletions in the genome as the original UCYN-A1 genome (54), but the DNA and amino acid
158   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

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

Mechanisms of UCYN-A symbiosis

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

Physical association

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

231 Hagino et al. (*52, 66*) were able to visualize *B. bigelowii* in Japan coastal waters because of the
232 presence of unique calcareous plates. TEM images showed 1-2 inclusions of what appears to be
233 UCYN-A (based on PCR amplification of parallel samples (*52*)), within the eukaryote cell (Fig.
234 3i, j, *66*)). These results are puzzling, since they contrast with FISH images that suggest one cell
235 (or several cells) attached to the surface of the eukaryote (*40, 41, 56*). Furthermore, in at least
236 some of the associations, the partners can easily be dislodged from each other by simple filtration
237 (*40*) which would not be true of an intracellular body. The contradictory results might be because
238 the calcified *B. bigelowii* is a distinct life stage (these organisms are known to have complex life
239 cycles) or because the different haptophyte species/strains have different mechanisms of
240 association (and different physical structures).

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

256   There are interesting differences between the diatom symbionts and UCYN-A. In UCYN-A,
257   photosystem I (PSI) is retained, whereas in the *Epithemia* symbionts both photosystems have
258   been lost. This was interpreted by Nakayama et al. (*63*) to mean that UCYN-A is less far on the
259   evolutionary trajectory to endosymbiosis, but it also could mean that PSI activity is advantageous
260   in the marine environment. Many bacteria in the oceans have proteorhodopsins, or anaerobic
261   anoxygenic photosynthetic pathways (*67*), suggesting that supplementation of energy
262   metabolism may be more important in the marine environment than the freshwater environments
263   where the diatom symbionts are found. UCYN-A is missing the phosphate sensor regulon
264   (*phoBR*), the nitrogen regulating protein PII (*glnB*) and any possible ABC-type
265   nitrate/sulfonate/bicarbonate transporter, relative to the *E. turgida* symbiont. In contrast, UCYN-
266   A has retained the Fe (iron) III transport genes (*afuABC*), which must be important in the Fe-
267   limited oligotrophic oceans. Interestingly, UCYN-A and the *E. turgida* spheroid body have
268   retained the NtcA transcription regulator. NtcA is a N-regulatory transcriptional activator that is
269   required for transcription of *nifH* genes in the absence of ammonium in at least some N_2-fixing
270   cyanobacteria (*68*). The lack of Amt transporters in the *E. turgida* spheroid body and UCYN-A
271   may force the *ntcA* gene to be constitutively up-regulated and be involved in stimulating N_2
272   fixation. It is also possible that NtcA in UCYN-A responds to the availability of carbon skeletons
273   (*69*), presumably supplied by the host. The metabolic similarities and differences between the
274   freshwater spheroid bodies and the marine planktonic UCYN-A must be critical for selection in
275   their unique habitats, to have survived such extreme genome reduction.

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

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

288 **UCYN-A and Evolution**

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

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

308 The diatom spheroid bodies are considered obligate endosymbionts, since they are inseparable
309 from the host cells and are synchronized and passed on to daughter cells during host cell division
310 (*65, 78*). In the case of the *Paulinella* chromatophore, it divides synchronously with the
311 amoeboid host suggesting a certain level of host-endosymbiont integration (*79*). It is still
312 unknown how UCYN-A attaches to the host or if the UCYN-A division is orchestrated by the
313 eukaryote (Fig. 4). However, a few micrographs have been observed that show two UCYN-A1
314 cells per host right before sunset, which suggests that the cyanobacteria UCYN-A divide prior to
315 the host (*42*).

316 The distinctions between "endosymbiont" and "organelle" have been debated for a long time.
317 One accepted characteristic used to distinguish between organelles (plastid) and endosymbionts
318 is the presence of a double membrane in organelles and whether proteins are imported (*75, 80*)
319 (Fig. 6). Most organellar proteins are encoded by nuclear genes and translated by host ribosomes,
320 in contrast to endosymbionts where all of the cytosolic proteins are encoded by the
321 endosymbiont genome. It has not yet been shown that the *Paulinella* chromatophore possesses a
322 sophisticated protein import apparatus, and it is debated whether it is an endosymbiont or an
323 organelle. Along with the unicellular cyanobacterial freshwater diatom symbionts, UCYN-A are
324 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

365   **Author contributions**

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

371   **Competing interest**

372   The authors declare no competing financial interests.

373
374   **References**
375
376   1.   Fowler, D. *et al*. The global nitrogen cycle in the twenty-first century. *Phil. Trans. R.*
377   *Soc. B*. **368**, 20130164 (2013).
378   2.   Raymond, J., Siefert, J. L., Staples, C. R. & Blankenship, R. E. The natural history of
379   nitrogen fixation. *Mol. Biol. Evol.* **21**, 541-554 (2004).
380   3.   Kneip, C., Lockhart, P., Voss, C. & Maier, U. G. Nitrogen fixation in eukaryotes-new
381   models for symbiosis. *BMC Evol. Biol.* **7**, 55 (2007).
382   4.   Geurts, R., Xiao, T. T. & Reinhold-Hurek, B. What does it take to evolve a nitrogen-
383   fixing endosymbiosis? *Trends Plant Sci.* **21**, 199-208 (2016).
384   5.   Vitousek, P. M. & Howarth R. W. Nitrogen limitation on land and in the sea: How can it
385   occur? *Biogeochem.* **13**, 87-115 (1991).
386   6.   Redfield, A. C. The biological control of chemical factors in the environment. *Am. Sci.*
387   **46**, 205-222 (1958).
388   7.   Ryther, J. & Dunstan, W. Nitrogen, phosphorus, and eutrophication in the coastal marine
389   environment. *Science* **171**, 1008-1013 (1971).
390   8.   Gruber, N. & Sarmiento, J. L. Global patterns of marine nitrogen fixation and
391   denitrification. *Global Biogeochem. Cy.* **11**, 235-266 (1997).
392   9.   Codispoti, L. A. An oceanic fixed nitrogen sink exceeding 400 Tg N a^{-1} vs the concept of
393   homeostasis in the fixed-nitrogen inventory. *Biogeosciences* **4**, 233-253 (2007).
394   10.   Capone, D., Zehr, J., Paerl, H. & Bergman, B. *Trichodesmium*, a globally significant
395   marine cyanobacterium. *Science* **276**, 1221-1229 (1997).
396   11.   Mague, T. H., Mague, F. C. & Holm-Hansen, O. Physiology and chemical composition
397   of nitrogen fixing phytoplankton in the central North Pacific Ocean. *Mar. Biol.* **41**, 213-
398   227 (1977).
399   12.   Villareal, T. A. Nitrogen-fixation by the cyanobacterial symbiont of the diatom genus
400   *Hemiaulus*. *Mar. Ecol. Prog. Ser.* **76**, 201-204 (1991).
401   13.   Villareal, T. A. Marine nitrogen-fixing diatom-cyanobacteria symbioses. In *Marine*
402   *Pelagic Cyanobacteria: Trichodesmium and other Diazotrophs*, E. J. Carpenter, D. G.

403   Capone, J. G. Rueter, Eds. (Kluwer Academic Publishers, The Netherlands, 1992), pp.
404   163-175.

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

407   15.   Foster, R. A., Carpenter, E. J. & Bergman, B. Unicellular cyanobionts in open ocean
408   dinoflagellates, radiolarians, and tintinnids: Ultrastructural characterization and immuno-
409   localization of phycoerythrin and nitrogenase. *J. Phycol.* **42**, 453-463 (2006).

410   16.   Zehr, J. & McReynolds, L. Use of degenerate oligonucleotides for amplification of the
411   *nifH* gene from the marine cyanobacterium *Trichodesmium thiebautii*. *Appl. Environ.*
412   *Microbiol.* **55**, 2522-2526 (1989).

413   17.   Paerl, H. W., Prufert, L. E. & Ambrose, W. W. Contemporaneous N_2 fixation and
414   oxygenic photosynthesis in the nonheterocystous mat-forming cyanobacterium *Lyngbya*
415   *aestuarii*. *Appl. Environ. Microbiol.* **57**, 3086-3092 (1991).

416   18.   Zehr, J., Mellon, M. & Zani, S. New nitrogen-fixing microorganisms detected in
417   oligotrophic oceans by amplification of nitrogenase (*nifH*) genes. *Appl. Environ.*
418   *Microbiol.* **64**, 3444-3450 (1998).

419   19.   Zehr, J. *et al.* Unicellular cyanobacteria fix N_2 in the subtropical North Pacific Ocean.
420   *Nature* **412**, 635-638 (2001).

421   20.   Karl, D. M., Church, M. J., Dore, J. E., Letelier, R. M. & Mahaffey, C. Predictable and
422   efficient carbon sequestration in the North Pacific Ocean supported by symbiotic nitrogen
423   fixation. *Proc. Natl. Acad. Sci. USA.* **109**, 1842-1849 (2012).

424   21.   Dore, J. E., Brum, J. R., Tupas, L. M. & Karl, D. M. Seasonal and interannual variability
425   in sources of nitrogen supporting export in the oligotrophic subtropical North Pacific
426   Ocean. *Limnol. Oceanogr.* **47**, 1595-1607 (2002).

427   22.   Montoya, J. P. *et al.* High rates of N_2 fixation by unicellular diazotrophs in the
428   oligotrophic Pacific Ocean. *Nature* **430**, 1027-1031 (2004).

429   23.   Grabowski, M. N. W., Church, M. J. & Karl, D. M. Nitrogen fixation rates and controls
430   at Stn ALOHA. *Aquat. Microb. Ecol.* **52**, 175-183 (2008).

431   24.   Falcón, L. I., Cipriano, F., Chistoserdov, A. Y. & Carpenter, E. J. Diversity of
432   diazotrophic unicellular cyanobacteria in the tropical North Atlantic Ocean. *Appl.*
433   *Environ. Microbiol.* **68**, 5760-5764 (2002).

434   25.   Mazard, S. L., Fuller, N. J., Orcutt, K. M., Bridle, O. & Scanlan, D. J. PCR analysis of
435   the distribution of unicellular cyanobacterial diazotrophs in the Arabian Sea. *Appl.*
436   *Environ. Microbiol.* **70**, 7355-7364 (2004).

437   26.   Church, M. J., Bjorkman, K. M., Karl, D. M., Saito, M. A. & Zehr, J. P. Regional
438   distributions of nitrogen-fixing bacteria in the Pacific Ocean. *Limnol. Oceanogr.* **53**, 63-
439   77 (2008).

440   27.   Langlois, R. J., Hummer, D. & LaRoche, J. Abundances and distributions of the
441   dominant *nifH* phylotypes in the Northern Atlantic Ocean. *Appl. Environ. Microbiol.* **74**,
442   1922-1931 (2008).

443   28.   Moisander, P. H. *et al.* Unicellular cyanobacterial distributions broaden the oceanic N_2
444   fixation domain. *Science* **327**, 1512-1514 (2010).

445   29.   Halm, H. *et al.* Heterotrophic organisms dominate nitrogen fixation in the South Pacific
446   Gyre. *ISME J.* **6**, 1238-1249 (2012).

447   30.   Agawin, N. S. R. *et al.* Dominance of unicellular cyanobacteria in the diazotrophic
448   community in the Atlantic Ocean. *Limnol. Oceanogr.* **59**, 623-637 (2014).

449 31. Bentzon-Tilia, M. *et al*. Significant N$_2$ fixation by heterotrophs, photoheterotrophs and
450   heterocystous cyanobacteria in two temperate estuaries. *ISME J*. **9**, 273-285 (2015).
451 32. Messer, L. F., Doubell, M., Jeffries, T. C., Brown, M. V. & Seymour, J. R. Prokaryotic
452   and diazotrophic population dynamics within a large oligotrophic inverse estuary. *Aquat.*
453   *Microb. Ecol*. **74**, 1-15 (2015).
454 33. Shiozaki, T., Nagata, T., Ijichi, M. & Furuya, K. Nitrogen fixation and the diazotroph
455   community in the temperate coastal region of the northwestern North Pacific.
456   *Biogeosciences* **12**, 4751-4764 (2015).
457 34. Luo, Y. W. *et al*. Database of diazotrophs in global ocean: abundance, biomass and
458   nitrogen fixation rates. *Earth Syst. Sci. Data* **4**, 47-73 (2012).
459 35. Farnelid, H., Turk-Kubo, K., Muñoz-Marín, MC. M. & Zehr, J. P. New insights into the
460   ecology of the globally significant uncultured nitrogen-fixing symbiont UCYN-A. *Aquat.*
461   *Microb. Ecol*. **77**, 125-138 (2016).
462 36. Biegala, I. C. & Raimbault, P. High abundance of diazotrophic picocyanobacteria (< 3
463   μm) in a Southwest Pacific coral lagoon. *Aquat. Microb. Ecol*. **51**, 45-53 (2008).
464 37. Bonnet, S., Biegala, I. C., Dutrieux, P., Slemons, L. O. & Capone, D. G., Nitrogen
465   fixation in the western equatorial Pacific: Rates, diazotrophic cyanobacterial size class
466   distribution, and biogeochemical significance. *Global Biogeochem. Cy*. **23** GB3012,
467   (2009).
468 38. Le Moal, M. & Biegala, I. C. Diazotrophic unicellular cyanobacteria in the northwestern
469   Mediterranean Sea: A seasonal cycle. *Limnol. Oceanogr*. **54**, 845-855 (2009).
470 39. Krupke, A. *et al*. In situ identification and N$_2$ and C fixation rates of uncultivated
471   cyanobacteria populations. *Syst. Appl. Microbiol*. **36**, 259-271 (2013).
472 40. Thompson, A. W. *et al*. Unicellular cyanobacterium symbiotic with a single-celled
473   eukaryotic alga. *Science* **337**, 1546-1550 (2012).
474 41. Krupke, A. *et al*. Distribution of a consortium between unicellular algae and the N$_2$ fixing
475   cyanobacterium UCYN-A in the North Atlantic Ocean. *Environ. Microbiol*. **16**, 3153-
476   3167 (2014).
477 42. Cabello, A. M. *et al*. Global distribution and vertical patterns of a prymnesiophyte-
478   cyanobacteria obligate symbiosis. *ISME J*. **10**, 693-706 (2016).
479 43. Cornejo-Castillo, F. M. *et al*. Cyanobacterial symbionts diverged in the late Cretaceous
480   towards lineage-specific nitrogen fixation factories in single-celled phytoplankton. *Nat.*
481   *Commun*. **7**, 11071 (2016).
482 44. Fay, P. Oxygen relations of nitrogen fixation in cyanobacteria. *Microbiol. Mol. Biol. Rev*.
483   **56**, 340-373 (1992).
484 45. Berman-Frank, I., Lundgren, P. & Falkowski, P. Nitrogen fixation and photosynthetic
485   oxygen evolution in cyanobacteria. *Res. Microbiol*. **154**, 157-164 (2003).
486 46. Church, M. J., Short, C. M., Jenkins, B. D., Karl, D. M. & Zehr, J. P. Temporal patterns
487   of nitrogenase gene (*nifH*) expression in the oligotrophic North Pacific Ocean. *Appl.*
488   *Environ. Microbiol*. **71**, 5362-5370 (2005).
489 47. Zehr, J. P. *et al*. Globally distributed uncultivated oceanic N$_2$-fixing cyanobacteria lack
490   oxygenic photosystem II. *Science* **322**, 1110-1112 (2008).
491 48. Tripp, H. J. *et al*. Metabolic streamlining in an open-ocean nitrogen-fixing
492   cyanobacterium. *Nature* **464**, 90-94 (2010).

493 49. Shi, X. L., Marie, D., Jardillier, L., Scanlan, D. J. & Vaulot, D. Groups without cultured
494 representatives dominate eukaryotic picophytoplankton in the oligotrophic South East
495 Pacific Ocean. *PLOS ONE* **4**, (2009).
496 50. Krupke, A. *et al.* The effect of nutrients on carbon and nitrogen fixation by the UCYN-
497 A–haptophyte symbiosis. *ISME J.* **7**, 1635-1647 (2015).
498 51. Needham, D. M. & Fuhrman, J. A. Pronounced daily succession of phytoplankton,
499 archaea and bacteria following a spring bloom. *Nat. Microbiol.* **1**, 16005 (2016).
500 52. Hagino, K., Onuma, R., Kawachi, M. & Horiguchi, T. Discovery of an endosymbiotic
501 nitrogen-fixing cyanobacterium UCYN-A in *Braarudosphaera bigelowii*
502 (Prymnesiophyceae). *PLOS ONE* **8**, e81749 (2013).
503 53. Thompson, A. *et al.* Genetic diversity of the unicellular nitrogen-fixing cyanobacteria
504 UCYN-A and its prymnesiophyte host. *Environ. Microbiol.* **16**, 3238-3249 (2014).
505 54. Bombar, D., Heller, P., Sanchez-Baracaldo, P., Carter, B. J. & Zehr, J. P. Comparative
506 genomics reveals surprising divergence of two closely related strains of uncultivated
507 UCYN-A cyanobacteria. *ISME J.* **8**, 2530-2542 (2014).
508 55. Zehr, J. How single cells work together: Are single-celled symbioses organelle evolution
509 in action? *Science* **349**, 1163-1164 (2015).
510 56. Martínez-Pérez, C. *et al.* The small unicellular diazotrophic symbiont, UCYN-A, is a key
511 player in the marine nitrogen cycle. *Nat. Microbiol.* **1**, 16163 (2016).
512 57. Carpenter, E. J. & Janson, S. Intracellular cyanobacterial symbionts in the marine diatom
513 *Climacodium frauenfeldianum* (Bacillariophyceae). *J. Phycol.* **36**, 540-544 (2000).
514 58. Bench, S. R., Heller, P., Frank, I. & Arciniega, M. Whole genome comparison of six
515 *Crocosphaera watsonii* strains with differing phenotypes. *J. Phycol.* **49**, 786-801 (2013).
516 59. Ran, L. *et al.* Genome erosion in a nitrogen-fixing vertically transmitted endosymbiotic
517 multicellular cyanobacterium. *PLOS ONE* **5**, e11486 (2010).
518 60. Villareal, T. A. Evaluation of nitrogen fixation in the diatom genus *Rhizosolenia* Ehr. in
519 the absence of its cyanobacterial symbiont *Richelia intracellularis* Schmidt. *J. Plankton*
520 *Res.* **9**, 965-971 (1987).
521 61. Villareal, T. A. Laboratory culture and preliminary characterization of the nitrogen-fixing
522 *Rhizosolenia-Richelia* symbiosis. *Mar. Ecol.* **11**, 117-132 (1990).
523 62. Nakayama, T. *et al.* Spheroid bodies in rhopalodiacean diatoms were derived from a
524 single endosymbiotic cyanobacterium. *J. Plant Res.* **124**, 93-97 (2011).
525 63. Nakayama, T. *et al.* Complete genome of a nonphotosynthetic cyanobacterium in a
526 diatom reveals recent adaptations to an intracellular lifestyle. *Proc. Natl. Acad. Sci. USA.*
527 **111**, 11407-11412 (2014).
528 64. Trapp, E. M., Adler, S. & Zauner, S. *Rhopalodia gibba* and its endosymbionts as a model
529 for early steps in a cyanobacterial primary endosymbiosis. *Endocytobiosis Cell Res.* **23**,
530 21-24 (2012).
531 65. Prechtl, J., Kneip, C., Lockhart, P., Wenderoth, K. & Maier, U. G. Intracellular spheroid
532 bodies of *Rhopalodia gibba* have nitrogen-fixing apparatus of cyanobacterial origin. *Mol.*
533 *Biol. Evol.* **21**, 1477-1481 (2004).
534 66. Hagino, K. *et al.* Extracellular calcification of *Braarudosphaera bigelowii* deduced from
535 electron microscopic observations of cell surface structure and elemental composition of
536 pentaliths. *Mar. Micropaleontol.* **125**, 85-94 (2016).
537 67. Kolber, Z. S. *et al.* Contribution of aerobic photoheterotrophic bacteria to the carbon
538 cycle in the ocean. *Science* **292**, 2492-2495 (2001).

539 68. Frías, J. E., Flores, E. & Herrero, A. Requirement of the regulatory protein NtcA for the
540   expression of nitrogen assimilation and heterocyst development genes in the
541   cyanobacterium *Anabaena* sp. PCC7120. *Mol. Microbiol.* **14**, 823-832 (1994).

542 69. Herrero, A., Muro-Pastor, A. M. & Flores, E. Nitrogen Control in Cyanobacteria. *J.*
543   *Bacteriol.* **183**, 411-425 (2001).

544 70. Adler, S., Trapp, E. M., Dede, C., Maier, U. G. & Zauner, S. *Rhopalodia gibba* and its
545   endosymbionts as a model for early steps in a cyanobacterial primary endosymbiosis. In
546   *Endosymbiosis*, 167-179, Springer Vienna, Vienna (2013).

547 71. Amin, S. A. *et al.* Interaction and signalling between a cosmopolitan phytoplankton and
548   associated bacteria. *Nature* **522**, 98-101 (2015).

549 72. Kazamia, E., Helliwell, K. E., Purton, S. & Smith, A. G. How mutualisms arise in
550   phytoplankton communities: building eco-evolutionary principles for aquatic microbes.
551   *Ecol. Lett.* **19**, 810-822 (2016).

552 73. Green, B. R. Chloroplast genomes of photosynthetic eukaryotes. *Plant J.* **66**, 34-44
553   (2011).

554 74. Dyall, S. D., Brown, M. T. & Johnson, P. J. Ancient invasions: from endosymbionts to
555   organelles. *Science* **304**, 253-257 (2004).

556 75. Nakayama, T. & Archibald, J. M. Evolving a photosynthetic organelle. *BMC Biol.* **10**, 35
557   (2012).

558 76. Nowack, E. C. *et al.* Endosymbiotic gene transfer and transcriptional regulation of
559   transferred genes in *Paulinella chromatophora*. *Mol. Biol. Evol.* **28**, 407-422 (2011).

560 77. Bodyl, A., Mackiewicz, P. & Gagat, P. Organelle evolution: *Paulinella* breaks a
561   paradigm. *Curr. Biol.* **22**, R304-306 (2012).

562 78. Miyagishima, S. Y. Mechanism of plastid division: from a bacterium to an organelle.
563   *Plant Physiol.* **155**, 1533-1544 (2011).

564 79. Melkonian, M. & Mollenhauer, D. Robert Lauterborn (1869-1952) and his *Paulinella*
565   chromatophora. *Protist* **156**, 253-262 (2005).

566 80. Theissen, U. & Martin, W. The difference between organelles and endosymbionts. *Curr.*
567   *Biol.* **16**, R1016-1017; author reply R1017-1018 (2006).

568 81. Hilton, J. A. *et al.* Genomic deletions disrupt nitrogen metabolism pathways of a
569   cyanobacterial diatom symbiont. *Nat. Commun.* **4**, 1767 (2013).

570 82. Church, M. J., Jenkins, B. D., Karl, D. M. & Zehr, J. P. Vertical distributions of nitrogen-
571   fixing phylotypes at Stn ALOHA in the oligotrophic North Pacific Ocean. *Aquat. Microb.*
572   *Ecol.* **38**, 3-14 (2005).

573 83. Needoba, J. A., Foster, R. A., Sakamoto, C. & Zehr, J. P. Nitrogen fixation by unicellular
574   diazotrophic cyanobacteria in the temperate oligotrophic North Pacific Ocean. *Limnol.*
575   *Oceanogr.* **52**: 1317-1327 (2007).

576 84. Foster, R. A. *et al.* Influence of the Amazon River plume on distributions of free-living
577   and symbiotic cyanobacteria in the western tropical north Atlantic Ocean. *Limnol.*
578   *Oceanogr.* **52**: 517-532 (2007).

579 85. Man-Aharonovich, D., Kress, N., Zeev, E. B., Berman-Frank, I. & Béjà, O. Molecular
580   ecology of *nifH* genes and transcripts in the eastern Mediterranean Sea. *Environ.*
581   *Microbiol.* **9**: 2354-2363 (2007).

582 86. Hashimoto, R., Yoshida, T., Kuno, S., Nishikawa, T. & Sako, Y. The first assessment of
583   cyanobacterial and diazotrophic diversities in the Japan Sea. *Fish. Sci.* **78**: 1293-1300
584   (2012).

585   87.   Brown, S. M. & Jenkins, B. D. Profiling gene expression to distinguish the likely active
586   diazotrophs from a sea of genetic potential in marine sediments. *Environ. Microbiol.* **16**,
587   3128-3142 (2014).
588   88   Bothe, H., Tripp, H. J. & Zehr, J.P. Unicellular cyanobacteria with a new mode of life:
589   the lack of photosynthetic oxygen evolution allows nitrogen fixation to proceed. *Arch.*
590   *Microbiol.* **192**, 783-790 (2010).
591   89.   Heller, P., Tripp, H. J., Turk-Kubo, K. & Zehr, J. P. ARBitrator: a software pipeline for
592   on-demand retrieval of auto-curated *nifH* sequences from GenBank. *Bioinformatics*
593   (2014).
594   90.   Messer, L. F. *et al.* High levels of heterogeneity in diazotroph diversity and activity
595   within a putative hotspot for marine nitrogen fixation. *ISME J.* **10**, 1499-1513 (2016).
596   91   Turk-Kubo, K. A. *et al.* Diazotroph community succession during the VAHINE
597   mesocosm experiment (New Caledonia lagoon). *Biogeosciences* **12**, 7435–7452 (2015).
598   92.   Pruesse, E., Peplies, J. & Glöckner, F. O. SINA: accurate high-throughput multiple
599   sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**, 1823-1829 (2012).
600   93   Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res.*
601   **32**, 1363-1371 (2004).
602   94.   Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local
603   alignment search tool. *J. Mol. Biol.* **215**, 403-410 (1990).
604   95.   Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools
605   for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**,
606   726-731 (2016).
607
608
609
610
611

Figure Legends

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

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

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

Figure 4. Methods used to discover the UCYN-A unicellular symbiosis. (**a**) Nitrogenase (*nifH*) genes amplified by PCR from DNA extracted from seawater samples. (**b**, **c**) Screening for UCYN-A in sorted populations of cells by screening for UCYN-A *nifH* by qPCR. (**d**) Genome
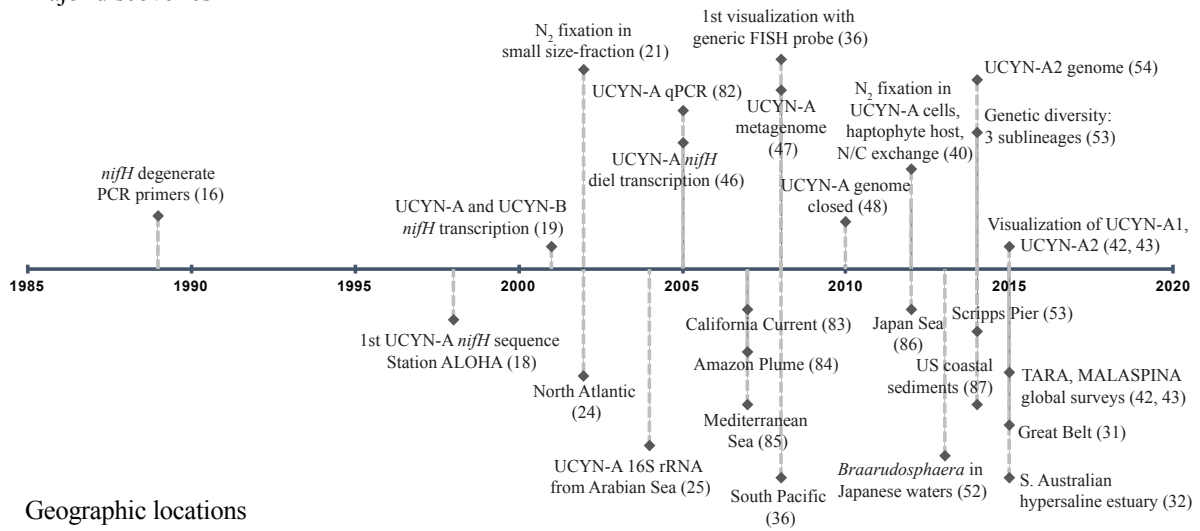
654 amplification and sequencing from sorted UCYN-A populations. (**e**) Single-cell
655 picophytoplankton cells sorted and screened for UCYN-A *nifH* by qPCR to identify host by 18S
656 rRNA gene PCR amplification. (**f**) Probes designed using the host 18S rRNA gene for
657 FISH/CARD-FISH methods to visualize cells. (**g**) Probes designed from the 16S rRNA gene
658 from genome sequence of UCYN-A for FISH/CARD-FISH methods. (**h**) Cells visualized in
659 seawater samples using FISH/CARD-FISH. Abbreviations: FACS – fluorescence activated cell
660 sorting; *nifH* – nitrogenase gene; PCR – polymerase chain reaction; qPCR – quantitative PCR;
661 FISH – Fluorescence *in situ* hybridization; and CARD-FISH – catalyzed reporter deposition
662 FISH.

663 **Figure 5**. Protein comparisons across genomes of UCYN-A1, the *Paulinella* chromatophore and
664 the *E. turgida* spheroid body. (**a**) Venn diagram showing pair-wise shared proteins and the 'core'
665 533 proteins that were identified as reciprocal match in all pair-wise BLASTP (*94*) searches. The
666 total number of proteins in each organism is noted on the Venn diagram in bold. (**b**) KEGG
667 Orthology (KO) assignment for the core proteins. (**c-e**) Barplots for individual genomes showing
668 KO distribution for proteins that are not in the core set. The genome files for UCYN-A1 and *E.*
669 *turgida* spheroid body were downloaded from the NCBI Genome database at
670 ftp.ncbi.nlm.nih.gov/genomes/all/ and the *Paulinella* chromatophore genome was downloaded
671 from the NCBI Nucleotide database. Protein fasta files were generated from Genbank files using
672 the BioPython script (gbk_to_faa.py). KO annotation for UCYN-A1 and *E. turgida* spheroid
673 body was downloaded from the KEGG website (http://www.genome.jp/kegg-bin/get_htext) and
674 htext was converted to tabulated files using a custom Perl script. KO for *Paulinella*
675 chromatophore was assigned using BlastKOALA (*95*).

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

Major discoveries

N₂ fixation in
small size-fraction (21)

1st visualization with
generic FISH probe (36)

UCYN-A qPCR (82)

UCYN-A
metagenome
(47)

UCYN-A2 genome (54)

N₂ fixation in
UCYN-A cells,
haptophyte host,
N/C exchange (40)

Genetic diversity:
3 sublineages (53)

nifH degenerate
PCR primers (16)

UCYN-A *nifH*
diel transcription (46)

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

1st UCYN-A *nifH* sequence
Station ALOHA (18)

California Current (83)

Japan Sea
(86)

Scripps Pier (53)

North Atlantic
(24)

Amazon Plume (84)

US coastal
sediments (87)

TARA, MALASPINA
global surveys (42, 43)

Mediterranean
Sea (85)

Great Belt (31)

UCYN-A 16S rRNA
from Arabian Sea (25)

South Pacific
(36)

Braarudosphaera in
Japanese waters (52)

S. Australian
hypersaline estuary (32)

Geographic locations

683

a

UCYN−A, ABZ80732.1
UCYN−A, ABZ80750.1
100 UCYN−A, ABZ80772.1
UCYN−A, ABZ80782.1
55 *Cyanothece* sp. CCY0110, EAZ92897.1
95 *Cyanothece* sp. ATCC 51142, AAW56988.1
90 *Gloeothece* sp. KO68DGA, BAF51663.1
Endosymbiont of *Rhopalodia gibba*, AAW57049.1
71 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)
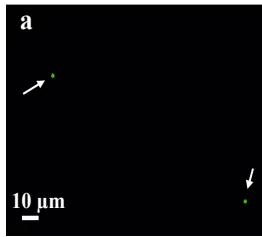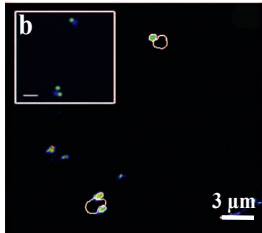Uncultured eukaryote clone SHAU528 (JQ222949.1)
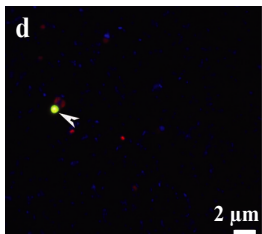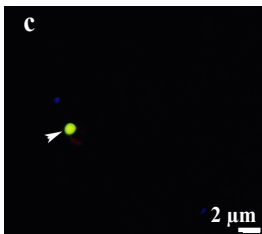
0.01

684

a

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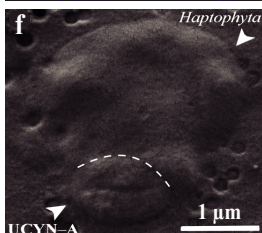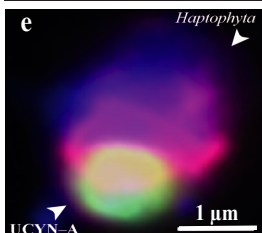
10 µm

b

2012: N$_2$ and CO$_2$ fixation rates in UCYN-A cells in association with a small Prymnesiophyte at Stn. ALOHA (40)

3 µm

c

d

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

2 µm

2 µm

e Haptophyta

f Haptophyta

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

UCYN–A 1 µm

UCYN–A 1 µm

g 2 µm

h

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

UCYN-A

prymnesiophyte 2 µm

1µm

1µm

i

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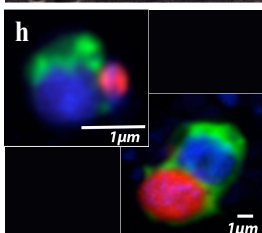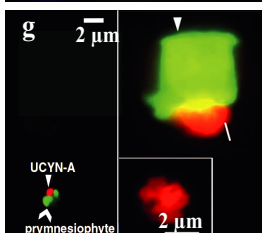
2016: SEM and TEM observations of cell surface structure of *B. bigelowii* in coastal areas (84)

1 µm

C

S

C

1 µm

k host

l host
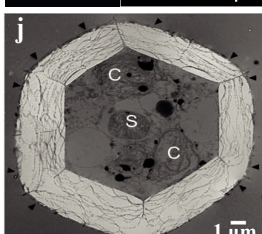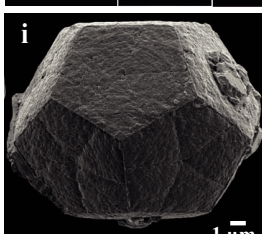
2016: N$_2$ and CO$_2$ fixation rates in two different UCYN-A associations in the North Atlantic (56)

UCYN-A 5 µm

UCYN-A 5 µm

685

Seawater
(mixed community)

FACS sorting

PCR

a

nifH gene

b

qPCR

16S/18S rRNA gene

FISH/ CARD-FISH

h

e

c

+

-

+

-

+

-

waste

Single sorted cells

Screening for
UCYN-A *nifH* (qPCR)
host 18S rRNA (PCR)

f

g

16S rRNA
gene

Bulk

d

Screening for
UCYN-A *nifH* by qPCR

Genome amplification
and sequencing

686

687

b **533 core proteins**

e **UCYN-A1 non-core proteins**

a

1132
85
48
867
168
533
452
118
617
1720

c *Paulinella* chromatophore non-core proteins

- Environ Information Processing
- Genetic Information Processing
- Metabolism

d *E. turgida* SB non-core proteins

- 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   2° Endosymbiosis

CB   1° PL   N   M   N   M

b   *Paulinella chromatophora*   c   *Epithemia turgida*

M   N

N

Chromatophore   Spheroid bodies

IM   OM   M

d   Possible models of symbiotic interactions in UCYN-A

IM   OM   UCYN-A   UCYN-A

UCYN-A   PL   N   PL   M

PL   N   PL   PL   N   PL

M   M

Haptophyte

689