Two subpopulations of *Crocosphaera watsonii* have distinct distributions in the North and South Pacific

Shellie R. Bench[1,2*], Ildiko Frank[1], Julie Robidart[1,3,4,5], Jonathan P. Zehr[1,3]

Affiliations:

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

[2]Current address: Stanford University, Department of Earth System Science, 473 Via Ortega, Rm 140 Stanford, CA 94305-4216

[3]Monterey Bay Aquarium Research Institute, 7700 Sandholdt Road, Moss Landing, CA 95039 USA

[4]Center for Microbial Oceanography: Research and Education, University of Hawaii, 1950 East-West Road, Honolulu, Hawaii 96822 USA

[5]Current address: Ocean Technology and Engineering Group, National Oceanography Centre, European Way, Southampton, SO14 3ZH, UK

[*]Corresponding author E-mail: shellierb@gmail.com Phone: 410-441-0055 Fax: 650-498-5099

Bench, Shellie R.; Frank, Ildiko; Robidart, Julie; Zehr, Jonathan P.. 2016 Two subpopulations of Crocosphaera watsonii have distinct distributions in the North and South Pacific. *Environmental Microbiology*, Early View.10.1111/1462-2920.13180

Running title: Natural abundances of two *Crocosphaera* sub-types

Abstract

Crocosphaera watsonii is a unicellular nitrogen (N_2)-fixing cyanobacterium with ecological importance in oligotrophic oceans. In cultivated strains there are two phenotypes of *C. watsonii* (large and small cell) with differences that could differentially impact biogeochemical processes. Recent work has shown the phenotypes diverged through loss or addition of type-specific genes in a fraction of their genomes, while the rest of the genomes were maintained at 99-100% DNA identity. Previous molecular assays for *C. watsonii* abundances targeted the conserved regions and therefore could not differentiate between phenotypes, so their relative distributions in natural communities were unknown. To determine phenotype distributions, this study developed and applied type-specific qPCR assays to samples from the North and South Pacific. Abundances of both *Crocosphaera* types declined sharply with depth between 45 and 75 m in both sites. In surface water small cells were 10 to 100 times more abundant than large cells in the N. Pacific; while in the S. Pacific the two phenotypes were nearly equal. Evidence for large cell aggregation was only found in N. Pacific samples. The differences in *C. watsonii* sub-populations in the North and South Pacific Ocean have direct implications for biogeochemistry and carbon export in oligotrophic gyres.

Introduction

Crocosphaera watsonii is a species of unicellular nitrogen (N_2)-fixing cyanobacteria ("UCYN" from here forward) that is important in marine primary production and biogeochemical cycling. This is especially true in oligotrophic waters where nitrogen is often a limiting nutrient, and *C. watsonii* is an important source of biologically available nitrogen (Zehr *et al.*, 2001; Falcon *et al.*, 2004; Montoya *et al.*, 2004; Kitajima *et al.*, 2009; Moisander *et al.*, 2010). UCYN are among the most abundant N_2-fixers in oceanic systems and measurements of UCYN abundance by direct microscopy counts and qPCR of the *nifH* gene are important for determining their contributions to the N cycle. Direct counts of UCYN abundances have ranged from 10^4 to 10^7 cells per liter in the North Pacific (Zehr *et al.*, 2001; Church *et al.*, 2005), and near 10^5 cells per liter in the Atlantic (Falcon *et al.*, 2004), and qPCR studies have reported between 10^3 and 10^6 gene copies per liter in multiple ocean basins (Zehr *et al.*, 2001; Falcon *et al.*, 2004; Church *et al.*, 2005; Church *et al.*, 2008; Langlois *et al.*, 2008; Moisander *et al.*, 2008; Moisander *et al.*, 2010). However, microscopic and qPCR enumerations of natural populations have treated all *Crocosphaera* cells as a single global population because the cells are morphologically similar by microscopy, and because there is a lack of genetic variation in the *nifH* gene.

The genetic conservation observed in *Crocosphaera nifH* sequences was also observed in sequences encoding 16S rRNA and a number of other genes that were examined in natural populations and cultivated strains (Zehr *et al.*, 2007). Despite their gene sequence conservation, two distinct phenotypic categories have been described in *C. watsonii* isolates. The first (large cell) phenotype has a cell-diameter of 4-6 μm, produces abundant extracellular polysaccharide (EPS), has higher photosynthetic efficiencies (F_v/F_m), and higher per-cell nitrogen fixation rates (Webb *et al.*, 2009; Sohm *et al.*, 2011). The other (small cell) type has a cell diameter less than 4 μm, and does not produce

noticeable amounts of EPS. There is also evidence that the small cell types grow in a narrower temperature range, and are missing some phosphorus scavenging genes that are found in the large cell types (Dyhrman and Haley, 2006; Webb *et al.*, 2009). More recently, genome comparisons showed that the large cell types contained a variety of genetic capabilities that were missing from the smaller genomes of the small cell types, such as EPS biosynthesis, iron stress response genes, and phosphorus metabolism genes (Bench *et al.*, 2011; Bench *et al.*, 2013).

Because of their genetic and metabolic differences, it is likely that the two phenotypes have different impacts on biogeochemical cycling. For example, EPS is a carbon-rich compound that can protect cells as well as cause aggregation which increases sinking rates (Passow *et al.*, 2001; Pereira *et al.*, 2009; Sohm *et al.*, 2011). As such, EPS production could make the large cell types greater contributors to carbon export from surface water than the small cell types. The differences in iron and phosphorus related genes further suggest that the two *C. watsonii* types may be differently adapted to their chemical environment, and therefore may have different niches. However, because previous methods used to measure natural *Crocosphaera* abundances were limited to viewing the species as a single population, very little is known about the distribution of the two types in the water column or the global oceans. In light of the ecologically relevant differences between types, it is important to determine their specific distributions in order to better understand the impact of the species on nutrient cycling and marine ecosystems. As a step towards that better understanding, this study was carried out with the following specific goals: 1) use recently identified phenotype-specific genes to develop specific and sensitive molecular assays to differentiate between *Crocosphaera* phenotypes and 2) apply those assays to water column samples from two regions in the Pacific Ocean in order to characterize differences in vertical and basin-wide distributions of the two *C. watsonii* types in natural populations.

Results

Recent genomic comparisons of multiple cultivated *Crocosphaera watsonii* strains identified gene sequences that were strain-specific and phenotype-specific (Bench *et al.*, 2011; Bench *et al.*, 2013). This enabled the design of the qPCR assays described in this study, which target four different genomic loci; two specific for large cell type strains, and two specific for small cell strains (Table 1). The previously used *Crocosphaera nifH* (also referred to as "Group B") primer probe sets (Moisander *et al.*, 2010) acted as the positive quantitative control during testing of each qPCR locus using DNA from six *C. watsonii* cultivated strains (see methods). All loci amplified quantitatively in reactions with DNA from strains with the targeted phenotype as expected, and did not amplify with DNA from non-target phenotype strains (Table S1). This was true for all loci, including reaction where DNA from a non-targeted phenotype strain (i.e. not expected to amplify) was ten-fold greater than DNA from the targeted strain.

After the qPCR assays were tested against individual strains, they were applied to samples from two research cruises (Fig. 1). There was good correlation across the N. Pacific samples between the two gene-specific assays within each phenotype (Fig. S1), and a paired t-test found no difference between the abundances reported by the assays ($p = 0.69$ for small cell assays, and $p = 0.71$ for large cell assays). The sum of the two phenotypes also correlated well with *nifH* abundances (Correlation coefficient = 0.9899, Fig. S2).

In the N. Pacific water column samples, total *nifH Crocosphaera* abundance was often $>10^6$ gene copies per liter in samples from 50 m and shallower, and declined by 2 to 3 orders of magnitude between 50 and 100 meters, decreasing to abundances of $\sim10^3$ gene copies per liter at and below 150 m (Fig. 2, upper panel). The small cell type accounted for the vast majority of total *Crocosphaera* at

depths shallower than 75 m, while the large cell type was more abundant in samples deeper than 100 m (Fig. 2). Small cell *Crocosphaera* abundances ranged from over 10^6 gene copies per liter at the surface to less than 10^3 copies per liter at and below 125 m (Fig. 2, middle panel). The abundance of the large cell types in the upper water column was 10^4 to 10^5 gene copies per liter, decreasing to ~10^3 copies per liter below 75 m (Fig. 2, lower panel). The ratio of small cell to large cell *Crocosphaera* (calculated by dividing the small cell gene copies by large cell gene copies for each sample) was typically between 10 and 1,000 in shallow samples, and between 0.1 and 1 in most samples below 75 m (Fig. S3).

Crocosphaera abundances were measured in sediment trap samples to assess possible differences in sinking processes between the two phenotypes. The ratio of the phenotypes in sediment traps was then compared with depth-integrated and spatially averaged values from water column samples collected during the same time (see methods). At 100 m, the average ratio of small cells to large cells was 51.4 in the water column and 42.9 in the sediment trap. At 150 m, the small: large ratio in the water column was 49.4 and the ratio in the sediment trap was 2.1.

For all water column samples, phenotype-specific assays were performed separately on the 10um and the 0.2 um filters that were collected in-line (see methods) and the relative contribution of the 0.2 µm filter to the total for each sample was calculated. For nearly all samples, close to 100% of small cells were found on the 0.2 µm filter (Fig. 3, upper panel and Fig. S4). Only two samples (out of 60) showed less than 60% of small cells on the 0.2 µm filter, and in the other 58 samples an average of 94% of small cell abundance was found on the 0.2 µm filter. In contrast, for many samples, especially in the first half of the cruise, less than 40% of the total large cell copies were found on the 0.2 µm filter (Fig. 3, lower panel and Fig. S5). In other samples (e.g. samples collected from below 75 m depth and

surface samples from the last two stations), nearly all large cells were found on the 0.2 µm filter, similar to the small cells.

A principal component analysis of environmental variables was used to investigate potential correlations between *Crocosphaera* abundances in BioLINCS samples and water column parameters. The first principal component (PC1) explained 48% of the observed variation, and was mainly related to depth, density, temperature, light, ammonia and nitrite. The second principal component (PC2) explained 30% of the variation and was mainly composed of oxygen, salinity, chlorophyll, phosphate and silicate. The two dimensional projection shows clustering of the water samples into three depth-related groups (Fig. 4): surface (orange symbols), chlorophyll maximum and adjacent depths (hereafter referred to as chl max, green symbols), and deeper water (blue symbols). In contrast to the chl max samples, the surface and deep samples showed little variation along PC1, but are well spread along PC2. Abundances of both *Crocosphaera* cell types (as well as total *Crocosphaera*) showed significant negative correlation with PC1.

The three clusters (surface, chl max, and deep) were also tested separately to identify correlations with environmental factors that were independent from depth and abundance (Table S2). In surface water samples, large cell abundances were positively correlated with phosphate and silicate and negatively correlated with salinity. In the chl max samples, both cell types were significantly positively correlated with temperature and PAR, and negatively correlated with salinity, density, chlorophyll, nitrate, and silicate. In addition, large cells had a negative correlation with ammonia. In the deep samples, no significant relationships were found between environmental variables and either cell type.

Abundances of the two *Crocosphaera* phenotypes were also measured in samples collected at three stations during a research cruise in the South Pacific Ocean in the austral fall of 2007. An in-

129 depth characterization of oceanographic conditions and total diazotroph population distributions are
130 described in Moisander *et al.* (2010). In these samples, abundances of both *Crocosphaera* types were
131 approximately 10^5 to 10^6 gene copies per liter near the surface, dropping to between 10^2 to 10^4 in
132 deeper water (Fig. 5, C). The ratio of large cells to small cells was relatively constant throughout the
133 water column, typically between 0.4 and 4 with an average of 1.5 for all samples (Fig. 5, D). Over 80%
134 of the total abundance of small cells was found on the 0.2 µm filter for all S. Pacific samples, with many
135 close to 100% and an overall average of nearly 93% (Table S3). Over 80% of the large cell types were
136 also captured on the 0.2 µm filter in all but three samples, with an average over 89% for all samples.
137

138 ## Discussion

139 Despite known phenotypic differences in cultivated strains, *Crocosphaera watsonii* has only
140 been assayed as a single population in the environment because of a lack of known genetic variation
141 and an assumption that it behaves ecologically as a single population. The qPCR assays described in
142 this study are necessary to examine natural populations for differences in the phenotypically distinct
143 groups of *C. watsonii*. Testing of the novel qPCR primers and probes demonstrated that all four qPCR
144 assays were robust with no evidence of cross-reactivity or inhibition from un-targeted *C. watsonii*
145 phenotypes. Strong agreement between the total *Crocosphaera nifH* abundances and the sum of the
146 two sub-types (Fig. S2) illustrated that these two types make up the entire natural *Crocosphaera*
147 community, at least for the samples in this study. As such, this study presents important details of *C.*
148 *watsonii* distributions in the Pacific Ocean, as well as new and robust tools that can be used to further
149 examine *C. watsonii* populations in other ocean basins and during other seasons and years.

In the upper water column N. Pacific samples, small cell *Crocosphaera* abundances and distributions were similar to total *nifH* abundances, and the large cell type *Crocosphaera* were much less abundant. However, in deeper water, large cell abundances were often much higher than the small cell abundances in the same samples. Intriguingly, while total *Crocosphaera* abundances were similar in surface water of the two locations, small cell types did not dominate the *Crocosphaera* populations in the S. Pacific. In fact, large cell abundances were on par with small cell abundances indicating that, at the time of sampling, conditions in the surface mixed layer of the S. Pacific were more favorable for the large cell type than conditions in the N. Pacific during the BioLINCS cruise (Fig. 5). However, at both locations, there was a slight dominance of large cell *Crocosphaera* below 75 m. In addition, the dominance of total *Crocosphaera* populations by the small cell population (i.e. the high ratio of small: large cells when integrating over the water column) was reduced in sediment trap samples relative to water column values collected in the N. Pacific, particularly at 150 m depth. The dominance of large cells in deeper water and the overrepresentation of large cell *Crocosphaera* in the sediment traps relative to the water column could be explained by any or all of the following; 1) faster sinking of the large cell type, 2) slower degradation of large cells during sinking, or 3) preferential grazing of the small cell type. All three mechanisms would be enhanced by EPS production in the large cells (Passow et al., 2001; Pereira et al., 2009; Sohm et al., 2011), emphasizing the importance of distinguishing the two types of *Crocosphaera* in the environment, since the two populations appear capable of playing different ecological roles. A previous study carried out microscopic cell counts to quantify relative abundance of two cell size classes of *Crocosphaera* in the western South Pacific (Webb *et al.*, 2009). In that study, the smaller cells were slightly more abundant than the larger cells, but populations were only examined within a narrow depth range near the surface (6-14 m), so deeper water patterns cannot be compared to the qPCR assays results of this study. Evidence of cell

aggregation was seen in the large cells (Webb *et al.*, 2009), which supports the possibility of differential export dynamics of the two phenotypes, and is different from the results presented here which did not show evidence of aggregation in the S. Pacific. Future experiments with additional environmental samples will be needed to fully understand the export processes and the conditions that favor the two phenotypes.

The principal component analyses carried out on the N. Pacific samples offer some clues about which environmental factors may affect *Crocosphaera* abundances. The largest component (PC1) was driven by factors that vary strongly with depth, and resulted in the samples clustering into three groups along the PC1 axis (Fig. 4). The statistically significant negative correlations with PC1 for both *Crocosphaera* phenotypes probably illustrate their observed depth-related decline as well as their lack of dependence on bioavailable nitrogen. In addition, the larger spread in PC1 for chl max samples was expected for samples that span the thermocline and contain the associated variability in depth-related factors (Robidart *et al.*, 2014). The second principal component (PC2) was driven by factors that varied more from station to station, (as opposed to depth) and resulted in a spreading of the surface and deep samples along this axis (Fig. 4).

The negative correlation with ammonia may be an indication of the type of N_2-fixation inhibition by ammonia that was observed in *Crocosphaera* culture experiments (Dekaezemacker, 2011; Garcia, 2014b) . Those studies found the effect of N_2-fixation inhibition was stronger under low light growth conditions (Garcia, 2014b), and also stronger in large cell strains compared to small cell strains (Dekaezemacker, 2011). Given those results, it is notable that there was a dominance of large cells in deeper water where ammonia concentrations are higher and light levels are lower. A factor that would favor the dominance of actively growing larger cells in deeper water is the observation that

195 certain small-cell strains will not grow under the same low light conditions where large cells area able
196 to grow (Garcia, 2013b).
197 In chl max samples (Fig. 4, green symbols), both cell types were significantly correlated with
198 salinity. As salinity is a conservative property of seawater, it appears that, within the chlorophyll max,
199 *Crocosphaera* distributions may be driven by mixing of disparate water types, rather than biological
200 interactions or nutrient variation. In surface samples, there was a correlation with phosphate and
201 silicate in large cells, and not small cells, indicating the small cell type may be less dependent on
202 phosphate. This is in contrast to genomic evidence (i.e. fewer copies of phosphorus related genes) that
203 suggests the small cell type is less adapted to low phosphorus conditions than the large cell strains
204 (Bench *et al.*, 2013). Of course, the small cell phenotype has a higher surface to volume ratio, so it may
205 also have a reduced need for phosphorus scavenging capabilities. There is also recent evidence of a
206 dramatic reduction in cell size in both phenotypes of cultured *Crocosphaera* under P and Fe co-
207 limitation (Jacq, 2014; Garcia 2014a). Those results underscore the importance of measuring both Fe
208 and P concentrations in the water where the two types are quantified as well as making microscopic
209 measurements of cell-size in natural populations. Future measurements of abundances of the two cell
210 types in other marine samples, particularly in the Atlantic where Fe is higher and P is lower than in the
211 Pacific, will help strengthen (or refute) the correlations observed in this study, and refine how
212 researchers understand *Crocosphaera* ecology.
213 Because all *C. watsonii* isolates have cell diameters that range from 3.5 µm to ~6 µm (Webb *et*
214 *al.*, 2009; Sohm *et al.*, 2011), it is expected that the vast majority of naturally occurring *C. watsonii* cells
215 should pass through the 10 µm filter and be captured on the 0.2 µm filter (which were arranged in-line
216 as described in the methods). However, as discussed above, the large cell phenotype is also known to
217 produce copious amount of EPS, and to form multi-cell aggregates as a result (Webb *et al.*, 2009; Sohm

218 *et al.*, 2011). Such aggregates could contribute to retention of the large cell type on the larger pore
219 size filter, as was observed in microscopic examination of the *Crocosphaera* community in the western
220 South Pacific (Webb *et al.*, 2009). In this study, the small cell type was found almost exclusively on the
221 0.2um filter in all samples in both ocean basins. However, in many of the N. Pacific samples, the
222 majority of large cells were found on 10um filters, supporting previous evidence of aggregation in the
223 large cell type.

224 The smaller fraction of large cells captured on 10um filters in the deeper N. Pacific samples
225 could be explained by aggregates breaking up as they sink, or by grazing. The release of cells by these
226 processes would result in a larger proportion passing through the 10 µm filter. This explanation would
227 mean that the presence of *C. watsonii* below 75 m is not indicative of those cells being active and
228 dividing (i.e. adapted to that depth), but rather a result of sinking and mixing. On the other hand, the
229 apparent lack of aggregation observed in large cell *Crocosphaera* in all S. Pacific samples (rather than
230 just deeper samples) suggests those cells were producing less EPS. There is some possibility that EPS
231 production within a genotype may be influenced by environmental factors. For example, a study of
232 Crocosphaera growth responses to inorganic nitrogen observed a plasticity in the C:N ratio of the large-
233 cell strain WH0003 (Dekaezemacker, 2011). While those authors did not observe changes in EPS
234 production, they suggested that modifying EPS production could be a mechanism that large cell
235 *Crocosphaera* could use to modulate their cellular C:N ratios. Further evidence from cultures and
236 genomic data strongly suggests that EPS production is an inherent, rather than inducible, trait in
237 *Crocosphaera* (Sohm *et al.*, 2011; Bench *et al.*, 2013). As such, the much lower incidence of capture on
238 the 10 µm in the S. Pacific surface samples suggests that basin may have had a different large cell sub-
239 type than the N. Pacific, where aggregation was more often observed. If that is the case, it will be
240 important to distinguish between the two large cell sub-types because a non-aggregating sub-type

241 would likely have export properties more similar to the small cell types. Distinct responses between
242 two small cell strains to changing CO_2 concentrations (Hutchins et al. 2013) provides evidence that
243 there are sub-types with different ecological adaptations within the larger two phenotypic categories
244 examined in this study. Additional genetic markers will be needed to design and carry out experiments
245 that could differentiate between the potential sub-types, and identify the physical and/or chemical
246 conditions that are more favorable to each type.
247 The three processes proposed above to explain the abundance patterns of the two size
248 fractions (aggregate break-down, grazing, and separate sup-types) have different predictions for the
249 metabolic state of the deeper population of cells. If the cells at those depths have simply sunk from
250 shallower water, they would not be expected to contribute fixed N to the deeper water where they
251 were observed. However, if they are a separate large-cell population that is adapted to those depths,
252 their contribution of new N to deeper water will need to be considered. Future experiments could
253 assess the biogeochemical contributions of the two sub-types using a variety of methods, including
254 measuring sinking rates, measuring N2-fixation rates in deep vs. shallow water, and by assessing the
255 viability of the deeper population through physiological fluorescence measurements and/or gene
256 expression levels of the two groups.
257

258 ## Conclusions

259 The qPCR assays developed for this study provide a novel method for quantifying two
260 phenotypes of *Crocosphaera* that were previously treated as a single, globally distributed population.
261 Resulting water column distribution patterns demonstrated, for the first time, that each phenotype has
262 a distinct biogeography, consistent with known phenotypic differences and suspected ecological

distinctions. Small cells were 10 to 1,000 times more abundant than large cells in the N. Pacific upper water column, but the two types were nearly equally represented throughout the water column in the S. Pacific. Furthermore, the evidence of large cell aggregation observed in the N. Pacific samples was not seen in the S. Pacific large cell populations. These patterns indicate that further sub-types of large cell *Crocosphaera* exist, with unique distributions the North and South Pacific, or alternatively, differences in physiology and/or grazing rates between the phenotypes result in differing degrees of aggregation in each basin. The observed patterns indicate that distinct controls determine the distributions of the two *Crocosphaera* phenotypes, many of which could vary over time as well as space. Because of the ecologically important differences between the two phenotypes demonstrated here, quantifying *C. watsonii* as two distinct groups is necessary for determining the global contribution of this keystone species to the carbon pump and marine biogeochemical cycles.

Experimental Procedures

Design and testing of qPCR assays

Previous comparisons of six *Crocosphaera watsonii* genomes identified genes unique to each phenotype (Bench *et al.*, 2011; Bench *et al.*, 2013). Two genes were chosen for each phenotype that were found in all genomes of one type, and were absent from all genomes of the other type. A primer-probe set was designed for each gene using Primer 3 (Rozen and Skaletsky, 1999) with a goal of obtaining a T_m of 64°C for primers, and 74°C for probes. All four genes and corresponding primer and probe sequences were used in nucleotide BLAST searches against the CAMERA (Sun *et al.*, 2011) and GenBank NT and WGS (Benson *et al.*, 2003) databases to verify that they did not have significant sequence similarity to other known organisms. The genes used for primer and probe design and the resulting sequences for all loci are listed in Table 1, with the design and testing of the *nifH* locus previously described (Moisander *et al.*, 2010). Dual-label probes were synthesized with FAM fluorescent tags and TAMRA quenchers. Reactions were set up in sterile PCR hoods using UV sterilized optical tubes or plates and contained 1.5 - 2 µl of template DNA plus 1 µl of each primer (10 µM), 0.5 µl of probe (10 µM), 12.5 µl TaqMan Gene Expression 2X Master Mix (Life Technologies, Grand Island, NY, USA), and water to a final volume of 25 µl. Amplification and detection was carried out on an ABI 7500 instrument using the following 2-step reaction: initial steps of 50°C for 2 minutes, then 95°C for 10 minutes, then 45 cycles of 90°C for 15 seconds, then 60°C for 60 seconds. Each run included 3 or 4 no template controls (NTCs) and a set of standards, in triplicate, with known gene copies from 10^0 to 10^7 per reaction. Following each run, the threshold cycle (C_t) values for each standard were plotted versus the log of its gene copy number to create a standard curve. The equation for that standard curve was used to calculate the gene copies in each of the sample reactions from the same run.

297 Standards were made from amplified genomic DNA from *C. watsonii* strains of the appropriate
298 phenotype. To avoid amplifying DNA from contaminants in non-axenic cultures, cells were sorted
299 using a flow cytometer prior to whole genome amplification (WGA) with Repli-g (Qiagen, Germantown,
300 MD, USA). The sorting and WGA were carried out as described in the methods used for genome
301 sequencing of *C. watsonii* strains (Bench *et al.*, 2011; Bench *et al.*, 2013). Amplified genomic DNA was
302 quantified using Pico Green (Life Technologies, Grand Island, NY, USA), and genome copies/µl were
303 calculated based on the DNA concentration and the genome sizes (Bench *et al.*, 2013). Appropriate
304 dilutions were made to generate a set of standards that contained 10^0 to 10^7 genome copies in 2 µl
305 (the volume used in each reaction). Multiple sets of the prepared genomic standards were compared
306 in triplicate to *nifH* linearized plasmid standards to verify the DNA quantification, and relative reaction
307 efficiency, and no significant differences were observed between the plasmid and any of the genomic
308 standards.
309 Tests for cross reactivity and inhibition were carried out for all loci using multiple mixtures of
310 DNA from different *C. watsonii* strains. Names and phenotypes of *C. watsonii* strains used are listed in
311 Table S1. Four test mixtures contained genomic DNA from WH8501 and WH0003 strains in the
312 following ratios: 1:3, 3:1, 1:10, and 10:1. Eight additional test mixtures contained WH8501 or WH0003
313 DNA mixed with 3-fold more DNA (final ratio of 1:3) from one of four additional strains (WH8502,
314 WH0401, WH0401, and WH0005). The primer-probe set for each locus was tested for amplification
315 and inhibition in triplicate qPCR reactions with the 12 different mixtures, which ranged over an order in
316 magnitude in target DNA and non-target DNA concentrations, and included samples that contained
317 only target DNA as well as only non-target DNA. Copy numbers from qPCR reactions were consistent
318 with DNA concentrations used in each reaction, and there was no amplification in any of the samples
319 that did not contain target DNA (Table S1). Because there was no observed cross-reactivity or

320 inhibition from non-target strains, all four primer-probe sets were determined to be appropriate for
321 use in environmental samples.
322
323 Sample collection, DNA extraction and qPCR of cruise samples
324 South Pacific samples were collected during the R/V Kilo Moana cruise KM0703 in March and
325 April of 2007. Cruise station locations as well as methods for water sample collection and processing
326 and DNA extraction were described previously (Moisander *et al.*, 2010). North Pacific samples were
327 collected at 8 stations during the BioLINCS cruise in September of 2011 just north of Station Aloha (Fig.
328 1). At each station, water samples were collected from multiple discrete depths (5, 25, 45, 75, 100,
329 125, 150, 175 m) with Niskin bottles mounted on a CTD rosette. Two to three liters of collected water
330 was filtered through two in-line Durapore filters (10 µm pore size, followed by 0.2 µm pore size).
331 Filters were placed in bead beater tubes with sterile glass beads, immediately flash-frozen in liquid
332 nitrogen and subsequently stored at -80°C until DNA was extracted. Sediment trap samples were
333 collected from seven depths (the three used in this study are 100, 150 and 500m) using a drifting
334 sediment trap that was deployed near Station 5 (Fig. 1). The trap drifted northeastward during the
335 BioLINCS cruise, on a track that was approximately 30º northeastward of the ship transit. In order to
336 relate qPCR abundances of ecotypes sampled from the ship's CTD Niskin bottles to samples collected in
337 the drifting sediment traps, samples exclusively from the ship's northeast trajectory were used. Details
338 of the sediment trap deployment, drift track and sample collection are described in Wilson *et al.*
339 (2014). Because the volume of sea water that contributed to the sediment trap samples is not known,
340 *Crocosphaera* abundances "per sample" are reported, and ratios of the two phenotypes (rather than
341 actual abundances) are used for comparisons between sediment trap samples and water column data.
342 In addition, water column abundances of each *Crocosphaera* type were integrated over equivalent

343 depths to the sediment traps (100 m, or 150 m), and the depth-integrated totals were used to calculate
344 the ratio of small to large cells at each station. The ratios were then averaged over the six stations that
345 approximated the multi-day drift track of the sediment trap in order to approximate the average water
346 column populations above the sediment trap during its deployment.

347 The DNA extraction protocol used for N. Pacific samples is a slight adaptation of the modified
348 DNeasy Plant MiniKit (Qiagen) protocol used to extract the S. Pacific samples (Moisander *et al.*, 2008;
349 Moisander *et al.*, 2010). Filters were thawed and 400 µl of AP1 buffer (provided in kit) was added to
350 each tube. Samples were subjected to three freeze-thaw cycles of rapid freezing in liquid N_2, followed
351 by rapid thawing in a 65°C heat block. The samples were then bead-beat in Mini-Beadbeater-96
352 (Biospec Inc.) for 2 minutes. Tubes were centrifuged briefly prior to addition of 45 µl (20 mg/ml) of
353 Proteinase K (Qiagen), vortexed briefly and incubated (with rocking) at 55°C for 1 hour. An RNaseA
354 digestion was then carried out by adding 4 µl of RNaseA to each sample, vortexing and incubating at
355 65°C for 10 minutes. The filters were removed from the tubes, and 130 µl of AP2 buffer (provided in
356 kit) was added to each tube followed by a brief vortex and a 10 minute incubation on ice. Tubes were
357 spun for 5 minutes at 14,000 RPM to pellet beads large precipitates, and the supernatant for each
358 sample was transferred to sterile 2 ml locking Sample tubes RB (Qiagen). DNA was extracted from the
359 transferred supernatant using the standard reagents and protocols for "Plant Cell & Tissues" with the
360 "DNeasy Plant Mini" kit in the QIAcube instrument (Qiagen). The final elution volume for each sample
361 was 100 µl.

362 The qPCR assays of environmental samples used the same reaction contents (except template
363 DNA), genomic DNA standards, and cycling conditions were as described for primer-probe testing
364 above. DNA extracts were diluted 1:5 (N. Pacific samples) or 1:1 (S. Pacific samples) and 1.5 µl of the
365 dilution was used in triplicate reactions. For the N. Pacific samples, *nifH* reactions contained 2 µl of

366 undiluted DNA extract in duplicate reactions. For the S. Pacific samples, previously determined *nifH*
367 abundance values (Moisander *et al.*, 2010) were used.
368

369 Acknowledgements

370 The authors would like to thank Pia Moisander, Kendra Turk-Kubo, and Ariel Rabines for
371 collecting samples and carrying out DNA extractions, as well as the captains and crew of the R/V Kilo
372 Moana during both research cruises. This work was funded by the NSF Center for Microbial
373 Oceanography: Research and Education (EF0424599) and by grants to JP Zehr from the Gordon and
374 Betty Moore Foundation.
375

376 References

377 Bench, S.R., Ilikchyan, I.N., Tripp, H.J., and Zehr, J.P. (2011) Two strains of *Crocosphaera watsonii* with highly
378 conserved genomes are distinguished by strain-specific features. Front Microbiol 2: 1-13.
379
380 Bench, S.R., Heller, P., Frank, I., Arciniega, M., Shilova, I.N., and Zehr, J.P. (2013) Whole genome comparison of
381 six *Crocosphaera watsonii* strains with differing phenotypes. J Phycol 49: 786-801.
382
383 Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. (2003) GenBank. Nucleic Acids Res
384 31: 23-27.
385
386 Church, M.J., Jenkins, B.D., Karl, D.M., and Zehr, J.P. (2005) Vertical distributions of nitrogen-fixing phylotypes at
387 Stn ALOHA in the oligotrophic North Pacific Ocean. Aquat Microb Ecol 38: 3-14.
388
389 Church, M.J., Bjorkman, K.M., Karl, D.M., Saito, M.A., and Zehr, J.P. (2008) Regional distributions of nitrogen-
390 fixing bacteria in the Pacific Ocean. Limnol Oceanogr 53: 63-77.
391
392 Dekaezemacker J, Bonnet S (2011) Sensitivity of N2 fixation to combined nitrogen forms (NO3- and NH4+) in two
393 strains of the marine diazotroph Crocosphaera watsonii (Cyanobacteria). Mar Ecol Prog Ser 438: 33–46. doi:
394 10.3354/meps09297
395
396 Dyhrman, S.T., and Haley, S.T. (2006) Phosphorus scavenging in the unicellular marine diazotroph *Crocosphaera*
397 *watsonii*. Appl Environ Microbiol 72: 1452-1458.
398

399 Falcon, L.I., Carpenter, E.J., Cipriano, F., Bergman, B., and Capone, D.G. (2004) N_2 Fixation by unicellular
400 bacterioplankton from the Atlantic and Pacific Oceans: phylogeny and in situ rates. Appl Environ Microbiol 70:
401 765-770.
402
403 Garcia, N.S., Fu, F-X, Breene, C.L., Yu, E.K., Bernhardt, P.W., Mulholland, M.R., Hutchins, D.A. 2013b. Combined
404 effects of CO2 and light on the unicellular N2-fixing cyanobacterium Crocosphaera watsonii: A comparison of
405 two isolates from the western tropical Atlantic Ocean. Eur. J. Phycol. 48:128-139.
406 doi:10.1080/09670262.2013.773383
407
408 Garcia, N.S., Fu, F., Sedwick, P.N., Hutchins, D.A. 2014a. Iron deficiency increases growth and nitrogen-fixation
409 rates in phosphorus-deficient marine cyanobacteria. ISME 9:238-245. doi:10.1038/ismej.2014.104
410
411 Garcia, N.S., Hutchins, D.A. 2014b. Light-limited growth rate modulates nitrate inhibition of dinitrogen fixation in
412 the marine cyanobacterium Crocosphaera watsonii. PLoS ONE 9: e114465. doi:10.1371/journal.pone.0114465
413
414 Hutchins et al. 2013. Taxon-specific response of marine nitrogen fixers to elevated carbon dioxide
415 concentrations. Nat Geosci. DOI: 10.1038/NGEO1858
416
417 Jacq V, Ridame C, Helguen SL, Kaczmar F, Saliot A. 2014. Response of the unicellular diazotrophic
418 cyanobacterium Crocosphaera watsonii to iron limitation. PLoS One 9: e86749.
419
420 Kitajima, S., Furuya, K., Hashihama, F., Takeda, S., and Kanda, J. (2009) Latitudinal distribution of diazotrophs and
421 their nitrogen fixation in the tropical and subtropical western North Pacific. Limnol Oceanogr 54: 537-547.
422
423 Langlois, R.J., Hummer, D., and LaRoche, J. (2008) Abundances and distributions of the dominant *nifH* phylotypes
424 in the Northern Atlantic Ocean. Appl Environ Microbiol 74: 1922-1931.
425
426 Moisander, P.H., Beinart, R.A., Voss, M., and Zehr, J.P. (2008) Diversity and abundance of diazotrophic
427 microorganisms in the South China Sea during intermonsoon. ISME J 2: 954-967.
428
429 Moisander, P.H., Beinart, R.A., Hewson, I., White, A.E., Johnson, K.S., Carlson, C.A. et al. (2010) Unicellular
430 cyanobacterial distributions broaden the oceanic N_2 fixation domain. Science 327: 1512-1514.
431
432 Montoya, J.P., Holl, C.M., Zehr, J.P., Hansen, A., Villareal, T.A., and Capone, D.G. (2004) High rates of N_2 fixation
433 by unicellular diazotrophs in the oligotrophic Pacific Ocean. Nature 430: 1027-1031.
434
435 Passow, U., Shipe, R.F., Murray, A., Pak, D.K., Brzezinski, M.A., and Alldredge, A.L. (2001) The origin of
436 transparent exopolymer particles (TEP) and their role in the sedimentation of particulate matter. Cont Shelf Res
437 21: 327-346.
438
439 Pereira, S., Zille, A., Micheletti, E., Moradas-Ferreira, P., Philippis, R.D., and Tamagnini, P. (2009) Complexity of
440 cyanobacterial exopolysaccharides: composition, structures, inducing factors and putative genes involved in
441 their biosynthesis and assembly. FEMS Microbiol Rev 33: 917-941.
442
443 Robidart, J.C., Church, M.J., Ryan, J.P., Ascani, F., Wilson, S.T., Bombar, D. et al. (2014) Ecogenomic sensor
444 reveals controls on N2-fixing microorganisms in the North Pacific Ocean. ISME J 8: 1175-1185.
445

Rozen, S., and Skaletsky, H. (1999) Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics Methods and Protocols*. Misener, S., and Krawetz, S.A. (eds). Totowa, NJ: Humana Press, pp. 365-386.

Sohm, J.A., Edwards, B.R., Wilson, B.G., and Webb, E.A. (2011) Constitutive extracellular polysaccharide (EPS) production by specific isolates of *Crocosphaera watsonii*. Front Microbiol 2.

Sun, S., Chen, J., Li, W., Altintas, I., Lin, A., Peltier, S. et al. (2011) Community cyberinfrastructure for Advanced Microbial Ecology Research and Analysis: the CAMERA resource. Nucleic Acids Res 39: D546-D551.

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

Wilson, S.T., del Valle, D.A., Segura-Noguera, M., and Karl, D.M. (2014) A role for nitrite in the production of nitrous oxide in the lower euphotic zone of the oligotrophic North Pacific Ocean. Deep Sea Research Part I: Oceanographic Research Papers 85: 47-55.

Zehr, J.P., Bench, S.R., Mondragon, E.A., McCarren, J., and DeLong, E.F. (2007) Low genomic diversity in tropical oceanic N_2-fixing cyanobacteria. P Natl Acad Sci USA 104: 17807-17812.

Zehr, J.P., Waterbury, J.B., Turner, P.J., Montoya, J.P., Omoregie, E., Steward, G.F. et al. (2001) Unicellular cyanobacteria fix N_2 in the subtropical North Pacific Ocean. Nature 412: 635-638.

Table and Figure legends

Table 1. Primer and probe sequences and gene source for qPCR assays. (All sequences area shown in 5' to 3' direction).

Figure 1. Sample locations and station numbers in the Pacific Ocean. South Pacific samples were collected in 2007, and North Pacific samples (details in inset) were collected in 2011.

Figure 2. Abundances for total (top panel) and two phenotypic sub-groups (middle and bottom) of *Crocosphaera watsonii* in samples collected during the BioLINCS cruise (station locations shown in Fig. 1). Gene copy numbers for all three assays are the sum of both size fractions for each sample, see supplemental material for plots of abundances of each size fraction.

Figure 3. Fraction of total abundance (i.e. the sum of both size fractions) found on 0.2μm filter for the small-cell phenotype of *Crocosphaera watsonii* (upper panel) and for the large-cell phenotype (lower panel) in samples collected during the BioLINCS cruise (station locations shown in Fig. 5, and abundances on each filter are shown in supplemental material (figures S4 and S5).

Figure 4. Principal component analysis of BioLINCS (N. Pacific) water samples. Symbols indicate depth where water sample was collected and are clustered into three depth-related groups, with surface samples in orange symbols, chl max in green, and deep samples in blue. Projections of environmental variables (red arrows) and *Crocosphaera watsonii* abundances (sum of both filters, purple arrows) are shown in the PC space multiplied by 10 and 5 respectively. PCA analysis did not include the depth as a variable. Total variance covered by the two components is 78% (48% by PC1 and 30% by PC2).

Figure 5. Abundances of two *Crocosphaera watsonii* phenotypes in North Pacific (A) and South Pacific (C) plotted according to the depth where each water sample was collected. The ratio of the two phenotypes (small:large) is also shown for each of the N. Pacific (B) and the S. Pacific (D) samples.