Genomic Characterization of Filamentous Cyanobacteria on Nantucket Beaches

Final Report to the Nantucket Biodiversity Initiative

Pia H. Moisander
Department of Biology, University of Massachusetts Dartmouth, 285 Old Westport Road, North Dartmouth, MA 02747, USA
Phone: 508-999-8222, email: pmoisander@umassd.edu

Abstract

A benthic, epiphytic cyanobacterium has been recently reported as invasive in temperate coastal beaches of the Nantucket Island, Massachusetts, USA, during late summer. Microscopic and molecular analyses were applied to identify the cyanobacterium and describe its phylogenetic position, and to characterize the microbial communities associated with it. The morphological observations in parallel with 16S rRNA and *nifH* results suggest that the cyanobacterium is closely related to *Hydrocoleum* sp. (Oscillatoriales, Phormidiaceae), with the closest previously described relative *Hydrocoleum lyngbyaceum*, and appears to represent a unique strain. The genus is a ubiquitous benthic cyanobacterium in tropical and subtropical waters and coral reefs, where it is at times considered a nuisance, while these observations in North Atlantic temperate waters may represent an expansion of its habitat range. Amplicon sequencing targeting the *nifH* gene confirmed that the cyanobacterium is capable of fixing N_2, and also hosts a relatively diverse community of *nifH*-gene containing bacteria typifying microbial mat diazotroph communities growing in association with it. *Hydrocoleum* was recently reported to be a potential producer of homoanatoxin-a, a neurotoxin, but genes for these toxins were not detected in the communities investigated in this study. Increasing sea surface temperatures in the North Atlantic Ocean may have contributed to this potential expansion of *Hydrocoleum* sp. habitat range.

Introduction

The goal of this study was to provide a first detailed characterization of a recently emerged benthic cyanobacterium reported from the temperate waters of New England, USA. Over the past few summers the cyanobacterium has formed considerable densities as epiphytic growth on seagrasses in the near shore areas of the Nantucket Island. Such growth can be considered a nuisance as the growth is shading the plants, and could have negative implications for boating and other recreational activities. Some benthic cyanobacteria are known to produce bioactive compounds including some that are characterized as toxins, and some of such cyanotoxins can be transferred to higher levels in the food web and are a concern for public health (Sivonen and Jones, 1999; Leao et al. 2012). Identifying the cyanobacterium and characterizing some of its functions is the first step toward determining its potential risks for public health and understanding reasons behind its recent apparent expansion.

The initial goal of this study was therefore to identify the cyanobacterium by using microscopic and molecular approaches. In parallel, the goal was to obtain initial information about functional aspects of the cyanobacterium by targeting several functional (protein-coding) genes. Since the cyanobacterium has not been isolated to pure culture, environmental samples were investigated in this study, thus the approaches had to consider community analyses. Studies of individual community members of marine benthic habitats such as microbial mats are complicated by the complex communities found in these systems, and high throughput approaches used in this study are powerful and necessary.

Use of the 16S rRNA gene in molecular characterization allows description of the entire microbial community present, providing information not only about the cyanobacterium but also on the bacterial communities living in association with it. In this study amplicon sequencing of the 16S rRNA gene using the Illumina MiSeq platform was used primarily to distinguish the cyanobacterium among the presumably numerous types of associated bacteria growing in these systems.

The cyanobacterium and associated communities were also characterized targeting a gene involved with the nitrogen (N_2) fixation process. Many benthic cyanobacteria are capable of fixing atmospheric N_2 using the nitrogenase enzyme, and the gene involved in the N_2 fixation reaction (*nifH* encoding for the dinitrogenase reductase protein in the nitrogenase enzyme), can be used to characterize the communities based on this function (Zehr et a. 2003). The presence

and phylogenetic affiliation of the *nifH* gene sequences from the samples, collected using a high throughput amplicon sequencing approach (Illumina MiSeq), was used as a second line of evidence for the phylogenetic affiliation of the cyanobacteria. The cyanobacterium could also live in association with other bacteria that are N_2 fixing, as is the case in many microbial mats (Moisander et al. 2006), and the approach used could provide detailed information of these communities as well.

Use of a high throughput approach results in thousands of sequence reads per sample, thus the samples are sequenced at a very high depth. Due to the high sequence yield, the proportion of each sequence in the set can be obtained at high confidence. Conclusions can thus be drawn about the phylogenetic affiliation of the dominant cyanobacterial sequence type (phylotype) in the set assuming that it originated from the filamentous cyanobacterium in question, forming a large portion of the DNA template in the sample. In addition to the high throughput approaches in this study, the cyanobacteria were initially investigated by amplifying the *ntcA* gene that is serving as a global nitrogen regulator only in cyanobacteria (Lindell and Post, 2001). Attempts were also made to amplify a gene involved with production of anatoxin-a/homoanatoxin-a, a toxin recently reported to be produced by similar benthic cyanobacteria in a tropical lagoon (Abed et al. 2006). Characterization of benthic cyanobacteria at the edges of their habitat range improves the understanding of factors contributing to their growth.

Materials and methods

Sampling and morphological observations

The samples were collected from the beach at a wading depth during low tide near the University of Massachusetts Boston Biological station (41.294N 70.040E) located on the north shore of the main island of Nantucket in the Nantucket Harbor. Small tufts of filaments were separated from the larger colonies growing epiphytically on seaweed using forceps cleaned with 70% ethanol, and were placed in sterile centrifuge tubes. Samples were kept on ice packs until arrival to the laboratory at the University of Massachusetts Dartmouth within a few hours. Smaller amounts of the filaments from these tufts were then removed with the sterile forceps into several bead beater tubes with approximately 0.1 g of sterile glass beads (approximately 1:1 volume of 0.1 and 0.5 mm diameter beads each). The tubes were then frozen at -80°C until further analysis. Samples were also saved in seawater as live, and preserved with 100% ethanol and Lugol solution for

microscopic analyses. For morphological observation, photographs were taken of the live samples using a Zeiss Axioplan epifluorescence microscope, using the DIC prism objectives at 10x and 40x (Figure 1).

DNA extraction

DNA was extracted using the Qiagen Plant minikit protocol with a few modifications. After addition of 400 µL lysis buffer from the kit (buffer AP1) and 4 µL RNase, the bead beater tubes were boiled at 100°C for 10 min. The tubes were agitated in a bead beater (BioSpec, Bartlesville OK) for 1 min at maximum speed, then kept at -80°C for 10 min or until frozen solid. The tubes were thawed at 55°C, then agitated and frozen again. The agitate-freeze-thaw cycle was repeated a total of three times. The samples were then processed following the kit protocol, ending with a 100-µL elution.

PCR conditions for 16S rRNA, nifH*,* ntcA*, and anaC*

For molecular characterization of the cyanobacteria, four different genes (16S rRNA, *nifH*, *ntcA*, *anaC*) were targeted using different PCR protocols (Table 1). For characterization of 16SrRNA gene diversity, PCR for high throughput amplicon sequencing was conducted using a nested protocol, to minimize the influence of non-target amplification on results. The PCR was conducted in triplicate, and each reaction contained 2.5 µL 10X buffer, 1.15 µL of 50 mM $MgCl_2$, 0.625 µL of 10 mM dNTPs, 0.125 µL of 100 µM Bact341F (Klindworth et al. 2013), 0.125 µL of 100 µM Bact785R, 0.1 µL of Platinum Taq polymerase (Life Technologies, Carlsbad, CA, USA), 0.375 µl BSA (20 mg mL^{-1}) and 1 µL of the DNA template, and were adjusted to 25 µL with nuclease free water. PCR conditions were as follows: initial 5 min at 95°C, then 25 cycles of 95°C for 40 s, 55°C for 2 min, 72° C for 1 min, and final extension at 72°C for 7 min, followed by a 4°C hold. For the second round PCR, the products from triplicates were pooled, and used as a template in the second round. Conditions were similar in the second round except the primers were replaced with the overhang 16S rRNA primers and only 5 cycles were included (Table 1). The triplicate products were pooled, then purified using the Ampure magnetic bead protocol (Agencourt, Beckman Coulter). Illumina indexes (Nextera XT index kit) were attached using a ligation PCR step, and purified again using the Ampure protocol. The products were quantified using Picogreen with a spectrofluorometer plate reader (Spectramax

M2, Molecular Devices), and samples combined at 4 nM concentration each in the sample that was multiplexed for sequencing.

Analysis of two independently processed samples were completed for the 16S rRNA using the amplicon sequencing protocol. The samples were included in a multiplexed sample sequenced at the Tufts University Core Facility using Illumina MiSeq v300 paired end method. The samples for this study were in a set of a total of 96 samples multiplexed (pooled with samples from other studies). Negative controls with nuclease free water added in place of DNA template were run in parallel with PCR and sequenced, although bands were not visible when the amplified products from these controls were run on the gel. The negatives were included for both 16S rRNA and *nifH* (see below).

The amplification of the gene encoding dinitrogenase reductase (*nifH*) was conducted by nested PCR protocol (Zehr and Turner 2001). The first round of nested amplification reactions included 2.5 µL 10X buffer, 1.25 µL of 50 mM $MgCl_2$, 0.5 µL of 10 mM dNTPs, 1 µL of 25 µM nifH3, 1 µL of 25 µM nifH 4, 0.11 µL of Platinum Taq polymerase (Life Technologies, Carlsbad, CA, USA), and 5 µL of the DNA template, and were adjusted to 26 µL with nuclease free water. The PCR conditions were an initial 3 min at 95°C, then 31 cycles of: 95°C for 30s, 57°C for 30 s, 72°C for 1 min, and 72°C for 7 min. The product from the first round was used as a template in the second round PCR reactions where overhang primers (nifH1_overhang and nifH2_overhang, Table 1) were used. The second round PCR reactions were otherwise identical to the first, except the template was one µL of the product from the first round reactions and the overhang primers were used, and only 25 cycles were included. The amplicons were processed using the Ampure magnetic bead protocol as described for 16S rRNA. All PCR reactions were done using a Bio-Rad C1000 thermocycler. One sample underwent the amplicon sequencing protocol. To account for potential reagent contamination (Zehr et al. 2003, Woyke et al. 2010), for both 16S rRNA and *nifH* amplicon sequencing, one PCR control sample was included in the reactions in which template was water instead of DNA, and the control was sequenced in parallel with samples.

For amplification of the *ntcA* gene, the PCR mix consisted of 2.5 µL 10X buffer, 0.8 µL 50 mM $MgCl_2$, 2.4 µL ntcA_1AF, and 0.8 µL ntcA 1AR (both primers at 25 µM stock concentration) (Lindell and Post 2001), 0.5 µL 10 mM dNTPs, 0.1 µl Taq polymerase (Thermo Fisher), and 16.9 µL nuclease free water. The amplification was done using the following

thermocycling program: initial 4 min at 94°C, then 30 cycles for 1 min at 94°C, 1 min at 55°C, and 2 min at 68°C, followed by 7 min at 72°C and a 4°C hold (ABI GeneAmp 2720 thermocycler). The products were separated on a 1.2% TAE gel electrophoresis, bands visualized with ethidinium bromide, then excised and purified (GeneJET Gel extraction kit, Thermo Scientific). The products were ligated and cloned using the pGEM-T vector system. Sequencing of five clones was conducted at the Massachusetts General Hospital (Cambridge, MA). Primers were trimmed from the sequences using the CLC Workbench (Cambridge, MA), and the sequences were then compared to the NCBI database using tblastx and blastn.

For amplification of the anatoxin-a/homoanatoxin-a gene *anaC*, the PCR mix consisted of 2.5 µL 10X buffer, 2-3 µL 25 mM $MgCl_2$, 0.5 µL anaC_GenF, and 0.5 µL anaC_GenR primers (both primers at 25 µM stock concentration, Table 1), 0.5 µL 10 mM dNTPs, 0.1 µl Taq polymerase (Thermo Fisher), and 2-3 µL DNA template. The reactions were adjusted to 25 µL with nuclease free water. The amplification was done using the following thermocycling program: initial 5 min at 94°C, then 30 cycles for 30 s at 94°C, 30-45 s at 50-55°C, and 30-45 s at 72°C, followed by 5-7 min at 72°C and a hold at 4°C (ABI GeneAmp 2720 thermocycler). The results were viewed by running the products on 1.5% TAE gel electrophoresis.

Analysis of the amplicon sequencing data (16S rRNA and nifH*)*

The paired end sequences were first combined and initial downstream analyses conducted in QIIME (Caporaso et al. 2010). For 16S rRNA, the paired sequences were binned against the Greengenes (http://greengenes.lbl.gov/cgi-bin/nph-index.cgi) database for determination of OTUs using the scripts in QIIME. The community composition was then assessed by proportion of total. Individual reads were subjected to Blastn to investigate DNA distance. For *nifH*, phylogenetic trees were made in Arb software (Ludwig et al. 2004) (database from Dr. Jonathan Zehr, University of California Santa Cruz; http://pmc.ucsc.edu/~wwwzehr/research/database/) comparing to reference sequences in the existing *nifH* database. In building the *nifH* tree with representative sequences, first 2000 sequences from the dataset were included, of which 66 were disarded due to poor translation.

Results

Morphology

Microscopy showed the colonies to be made of long non-heterocystous, non-branching filaments (Figure 1). The cells were disc shaped and thinner than wide with rounded ends, and the filaments had a relatively thin sheath. Hormogonia and some variation in filament width were observed (Figure 1).

Identification and community composition based on the 16S rRNA gene

Several thousand sequences were obtained for both samples processed with the 16S rRNA community analysis. The 16S rRNA sequences represented a diverse community. The most abundant bacterial class was Alphaproteobacteria, at almost 40% of all sequences. Altogether, Proteobacteria formed 45-48% of the sequences (Figure 2A). The next most abundant phylum was Bacteroidetes, at 23-27% of sequences (Flavobacteria forming the majority), and cyanobacterial sequences were the third most abundant phylum.

In genus-level binning of the sequences against the Greengenes 16S rRNA database, the dominant cyanobacterium in the community was most closely related with *Hydrocoleum* sp. (Cyanobacteria, Oscillatoriophycideae, Oscillatoriales, Phormidiaceae) (Figure 2B). This genus formed 13.9 and 18.7% of all 16S rRNA sequences in the two samples investigated, and 75-77% of all cyanobacterial 16S rRNA sequences. The second most common cyanobacteria related sequence was from eukaryotic chloroplasts and related to algae within the Stramenopiles order (8.2-11% of all cyanobacterial sequences); this order includes common marine phytoplankton such as diatoms. The third most common cyanobacterial sequence was identified as *Halomicronema* sp. (Synechococcophycideae, Pseudanabaenales, Pseudanabaenaceae), and formed 5.6-10% of the cyanobacterial sequences.

A representative 16S rRNA sequence (sequence 21530:13889 in the dataset) corresponding to the *Hydrocoleum* cluster of sequences had a 99% nucleotide identity with *Hydrocoleum glutinosum* M2 (GU238270) from the Indian Ocean (Palinska et al. 2015). It had a 97% identity with *Trichodesmium erythraeum* IMS101 strain genome (NR_074275).

Identification and community composition based on the nifH *gene*

A total of 66507 *nifH* sequences were obtained after the ends were paired. One major cyanobacterial sequence was present in the dataset (Figure 3). This phylotype formed 429 of 1934 randomly chosen sequences from the *nifH* dataset, thus roughly 22% of the *nifH* sequences. The best match in the database and to which the sequence had an approximately 97% identity with at the amino acid level (~95% at the nucleotide level) was *Hydrocoleum lyngbyaceum* (tblastx) (accession number GU238282.1), a sequence obtained from the Adriatic Sea (Mediterranean Sea). Other reported *Hydrocoleum* spp. *nifH* sequences available in the database clustered with the sequences from this study at 92-94% amino acid identity. Sequences from this study had an approximately 93% identity with *Trichodesmium thiebautii* and 95% identity with the *T. erythraeum* IMS 101 genome (amino acid level). The representative *nifH* sequence had 94% nucleotide identity with the benthic cyanobacteria *Lyngbya majuscula* (accession number AY115593.1) and *Okeania hirsuta* (accession number KC992990.1).

Many other *nifH* sequences were present in the community, and most of them were closely related with previously reported uncultivated marine bacteria (Figure 4). The phylogenetic tree with these representative phylotypes demonstrates the presence of primarily one cyanobacterial *nifH* sequence type in the sequence set, although a few other low abundance phylotypes were observed (not shown). Several *nifH* phylotypes clustered with Gammaproteobacteria, and included sequences closely related with *Pseudomonas stutzeri* (Figure 4). The non-cyanobacterial sequences included many representatives within Cluster 3, which is a cluster with relatively long branches, thus large differences among phylotypes, and includes many anaerobes. Many of the non-cyanobacterial sequences from this study clustered near uncultivated bacterial sequences previously recovered from microbial mats, saltmarshes, and rhizosphere environments (Figure 4). Many of the uncultivated phylotypes in this cluster are distant to any cultivated bacteria (Zehr et al. 2003).

The ntcA *gene as an indicator of phylogenetic placement*

Five clones were sequenced (using Sanger sequencing) of the successful amplification of the *ntcA* gene, and three of these produced high quality sequence data that were very similar among the three sequences. The best matches based on blastn (nucleotide comparison) to all of these clones (M5312D03, M5312D05, and M5312D06) was the global nitrogen regulator protein NtcA

in the cyanobacterium *Cyanothece* sp. PCC 7425 (Oscillatoriophycidae, Chroococcales). Although this was the closest match in GenBank, the sequence was quite distant at the nucleotide level, with only a 76% identity (100% query coverage, maximum e-value 2e-80). Comparison of the translated nucleotide to the translated nucleotides in NCBI database (tblastx) resulted in *Cylindrospermum stagnale* PCC7417 (accession CP003642.1) (Nostocales) or *Nostoc* sp. PCC 7170 (CP003548.1) (Nostocales) as the best matches (each at 90% identity). Several other filamentous and coccoid cyanobacterial genomes had 87-90% identity with the sequence based on tblastx.

Amplification of the anatoxin-a/homoanatoxin-a gene anaC

The anatoxin-a gene did not amplify from the samples despite several attempts to optimize the PCR protocol.

Discussion

The lack of heterocysts, lack of branching, and the cell shape all indicate that the cyanobacterium belongs to Oscillatoriaceae. The thin sheath suggests it is not a *Lyngbya* sp.

The overall results from 16S rRNA and *nifH* strongly suggest the filamentous 'alga' also called the 'witches hair' growing as an epiphyte on seagrasses on the Nantucket shorelines is indeed a cyanobacterium and belongs to the *Hydrocoleum* genus. This genus is a common cyanobacterium in tropical and subtropical waters, but not generally discussed in the context of temperate waters; in fact this may be the first such report.

Based on 16S rRNA, *nifH* and phycobilisome genes, *Hydrocoleum* collected from tropical New Caledonian waters was suggested to be closely related with the planktonic, globally important N_2-fixing cyanobacterium *Trichodesmium* (Abed et al. 2006). Morphological and phylogenetic analyses with a large number of strains confirmed this observation (Palinska et al. 2015). In this study, there was a tight clustering of *Hydrocoleum* from the populations on Nantucket Island sampled in this study and *Trichodesmium* sequences based on the *nifH* gene, and they were also closely related based on the 16S rRNA gene. Notably, however, the 16S rRNA gene was very closely related with a number of other benthic cyanobacteria, yet the best consistent match was to *Hydrocoleum* from the large dataset.

Nitrogen fixation occurring at night is known to be present in *Hydrocoleum*, and it is thought to be an important contributor to N_2 fixation in certain subtropical microbial mats and coral reef environments (Yu et al. 2011, Palinska et al. 2015). While the closest relative of the *nifH* sequences is *H. lyngbyaceum*, the distance of 97% at amino acid and 95% at nucleotide level of the dominant *nifH* sequence are still relatively distant, suggesting that the strain is unique, and could have specific adaptations to local conditions. However, there are relatively few *Hydrocoleum* spp. *nifH* sequences in the public databases.

The *ntcA* sequence phylotype recovered in this study was also very distant from published sequences, and if originated from *Hydrocoleum* they would suggest the strain is unique. However, more clones for *ntcA* would confirm the fact that the sequences obtained were from *Hydrocoleum* sp., and not from the associated cyanobacteria that were observed at lower abundances in both the 16S rRNA and *nifH* sequence data.

A previous study suggested that mats containing *Hydrocoleum* may have been homoanatoxin-a producing (Mejean et al. 2010), and methodological approaches were recently published on targeting genes involved in this pathway in cyanobacteria (Rantala-Ylinen et al. 2011). However, several PCR attempts, including protocols with reduced stringency in this study failed to show presence of genes involved with anatoxin-a or homoanatoxin-a production, suggesting the samples studied did not have producers of these toxins present. The aspect of production of bioactive compounds and potential toxins by these cyanobacteria should merit future study.

To the best of our knowledge, the habitat range of *Hydrocoleum* is currently considered to include only tropical and subtropical oceans, while this may be the first reported observation in temperate waters (Abed et al. 2006, Villeneuve et al. 2012). The observation of *Hydrocoleum* in the North Atlantic at 41°N could be indicative of its expanding habitat range to more northern latitudes in the warming oceans. Within the seasonal and interdecadal variability, an overall increase is occurring in both summertime and wintertime average sea surface temperatures in the North Atlantic (Figure 5). Such increase could have contributed to the increasing habitat range of this cyanobacterium previously considered to be a subtropical and tropical genus.

Acknowledgements

I thank M. Daley for assistance with molecular analyses, and S. Oktay (University of Massachusetts Boston) and T. Hett (Brown University) for assistance with sampling. Funding from the Nantucket Shellfish Association (through the Nantucket Biodiversity Initiative), and the University of Massachusetts Dartmouth supported this study.

Figure legends

Figure 1. Microscopic images of live samples collected from Nantucket in August 2014.

Figure 2. A. 16S rRNA community in two independent samples of the cyanobacteria based on amplicon sequencing. B. Proportion of 16S rRNA sequences in different cyanobacterial groups, identified to the genus level or the next classification level possible. p_, phylum; c_, class; o_, order; f_, family; g_, genus.

Figure 3. Phylogenetic (neighbor-joining) tree for the dominant cyanobacterial sequence in the *nifH* sequence set. First 2000 sequences were used from the sequence set (a total of >66,000 sequences).

Figure 4. Phylogenetic (neighbor-joining) tree of sequences representing some of the major phylotypes in the *nifH* sequence set. The major phylotypes were determined based on the first 2000 sequences in the dataset examined in Arb. This tree does not show repeat sequences for these phylotypes. Sequences from this study start with 'Hwi'.

Figure 5. Mean sea surface temperature in the North Atlantic in February (A) and August (B) between 1856 and 2014 based on ship observations and remote sensing data from the National Oceanographic and Atmospheric Administration (http://www.esrl.noaa.gov/psd/data/timeseries/AMO/, Kaplan et al. 1998).

Literature Cited

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI et al (2010) QIIME allows analysis of high -throughput community sequencing data. *Nature Methods* 7:335-336

Kaplan A, Cane M, Kushnir Y, Clement A, Blumenthal M, and Rajagopalan B (1998) Analyses of global sea surface temperature 1856-1991. *J Geophys Res* 103:18567-18589

Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, and Gloeckner FO (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* 41.

Leao PN, Engene N, Antunes A, Gerwick WH, Vasconcelos V (2012) The chemical ecology of cyanobacteria. *Natural Product Reports* 29:372-391

Lindell D, and Post AF (2001) Ecological aspects of *ntcA* gene expression and its use as an indicator of the nitrogen status of marine *Synechococcus* spp. *Appl Environ Microbiol* 67: 3340-3349.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, Buchner A, Lai T, Steppi S, Jobb G, Forster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann S, Hermann S, Jost R, Konig A, Liss T, Lussmann R, May M, Nonhoff B, Reichel B, Strehlow R, Stamatakis A, Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A, and Schleifer KH (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* 32: 1363-1371.

Mejean A, Peyraud-Thomas C, Kerbrat AS, Golubic S, Pauillac S, Chinain M, Laurent D (2010) First identification of the neurotoxin homoanatoxin-a from mats of *Hydrocoleum lyngbyaceum* (marine cyanobacterium) possibly linked to giant clam poisoning in New Caledonia. Toxicon 56:829-835

Moisander PH, Shiue L, Steward GF, Jenkins BD, Bebout BM, Zehr JP (2006) Application of a *nifH* oligonucleotide microarray for profiling diversity of N2-fixing microorganisms in marine microbial mats. *Environ Microbiol* 8:1721-1735

Palinska KA, Abed RMM, Charpy L, Langlade M-J, Beltran-Magos Y, Golubic S (2015)

Morphological, genetic and physiological characterization of *Hydrocoleum*, the most common benthic cyanobacterium in tropical oceans. *Eur J Phycol* 50:139-154

Rantala-Ylinen A, Känä S, Rouhiainen L, Wahlsten M, Rizzi E, Berg K, Gugger M, Sivonen K. 2012. Anatoxin-a synthetase gene cluster of the cyanobacterium *Anabaena* sp. strain 37 and molecular methods to detect potential producers. Appl *Environ Microbiol* 77:7271-7278

Sivonen K, Jones G (1999) Cyanobacterial toxins. In: Chorus I, Bartram, J. (ed.) Toxic cyanobacteria in water. E&FN Spon, pp

Villeneuve A, Laurent D, Chinain M, Gugger M, Humbert JF (2012) Molecular characterization of the diversity and potential toxicity of cyanobacterial mats in two tropical lagoons in the South Pacific Ocean. *J Phycol* 48:275-284

Woyke T, Tighe D, Mavromatis K, Clum A, Copeland A, Schackwitz W, Lapidus A, Wu DY, McCutcheon JP, McDonald BR, Moran NA, Bristow J, and Cheng JF (2010) One Bacterial Cell, One Complete Genome. *PLoS One* 5.

Yu J, and Brand, JJ (2011) *Hydrocoleum* is primarily responsible for nitrogen fixation in Texas Gulf coast microbial mats. *J Phycol* 47: S9-10

Zehr JP, and Turner PJ (2001) Nitrogen fixation: Nitrogenase genes and gene expression. *Methods in Microbiology*, Vol 30 30: 271-286.

Zehr JP, Crumbliss LL, Church MJ, Omoregie EO, and Jenkins BD (2003) Nitrogenase genes in PCR and RT-PCR reagents: implications for studies of diversity of functional genes. *Biotechniques* 35: 996-+.

Zehr JP, Jenkins BD, Short SM, Steward GF (2003) Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Environ Microbiol* 5:539-554

Table 1. Primers used in this study. F, Forward primer; R, Reverse primer.

Gene	Gene function		Primer name: Primer sequence	Reference
16S rRNA	Microbial small subunit ribosomal RNA	F R F R	S-D-Bact-0341-b-S-17: 5'-CCTACGGGNGGCWGCAG-3' S-D-Bact-0785-a-A-21: 5'-GACTACHVGGGTATCTA ATCC-3' 341_overhang: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG-3' 785_overhang: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC-3'	Klindworth et al. 2013
anaC	Proline adenylation in the Homoanatoxin-a production pathway	F R	anaC-genF: 5'-TCTGGTATTCAGTCCCCTCTAT-3' anaC-genR: 5'-CCCAATAGCCTGTCATCAA-3'	Rantala-Ylinen et al. 2011
nifH	Dinitrogenase reductase protein in the nitrogenase enzyme	F R F R	nifH1_overhang: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGYGAYCCNAARGCNGA-3' nifH2_overhang:5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGADNGCCATCATYTCNCC-3' nifH4: TTYTAYGGNAARGGNGG nifH3: ATRTTRTTNGCNGCRTA	This study (based on Zehr and Turner 2001 and Illumina, Inc. 2014) Zehr and Turner 2001
ntcA	Global nitrogen regulator	F R	1AF: 5'-ATH TTY TTY CCB GGG GAY CCD GC-3' 1AR: 5'-ATG GCY TCG GCK ATG GCY TGR T-3'	Lindell and Post 2001

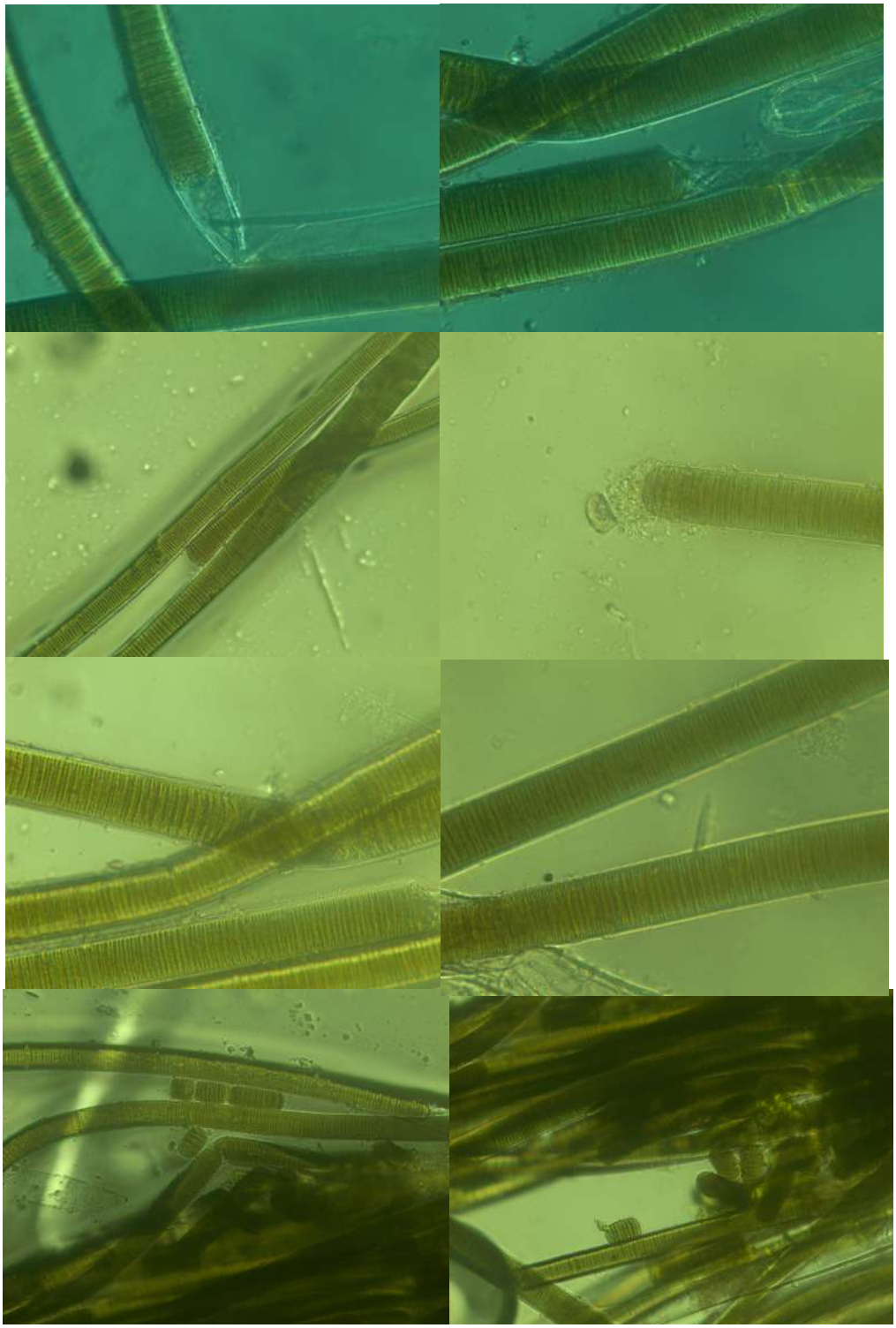

Figure 1.

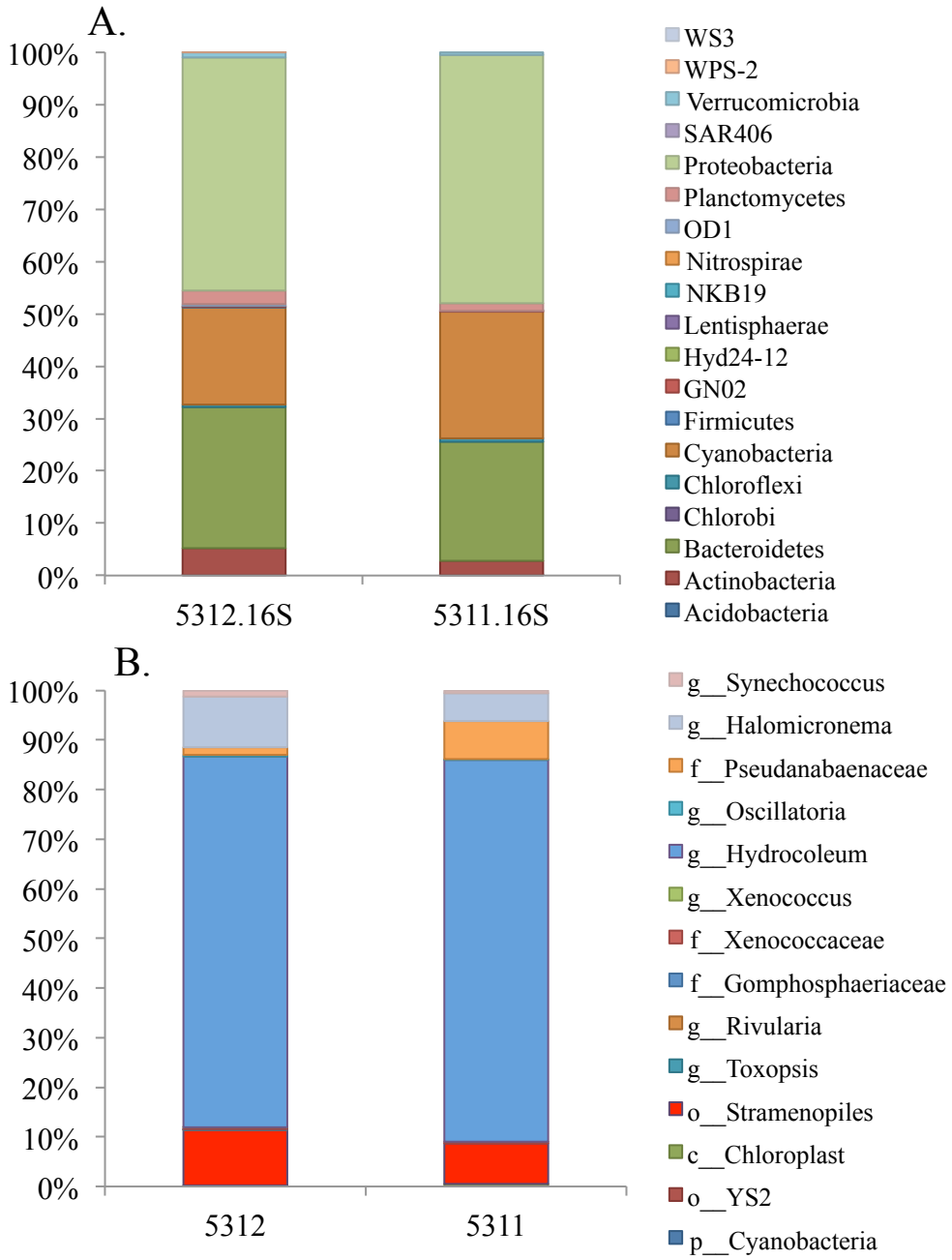
A.
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
5312.16S
5311.16S
WS3
WPS-2
Verrucomicrobia
SAR406
Proteobacteria
Planctomycetes
OD1
Nitrospirae
NKB19
Lentisphaerae
Hyd24-12
GN02
Firmicutes
Cyanobacteria
Chloroflexi
Chlorobi
Bacteroidetes
Actinobacteria
Acidobacteria
B.
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
5312
5311
g__Synechococcus
g__Halomicronema
f__Pseudanabaenaceae
g__Oscillatoria
g__Hydrocoleum
g__Xenococcus
f__Xenococcaceae
f__Gomphosphaeriaceae
g__Rivularia
g__Toxopsis
o__Stramenopiles
c__Chloroplast
o__YS2
p__Cyanobacteria

Figure 2.

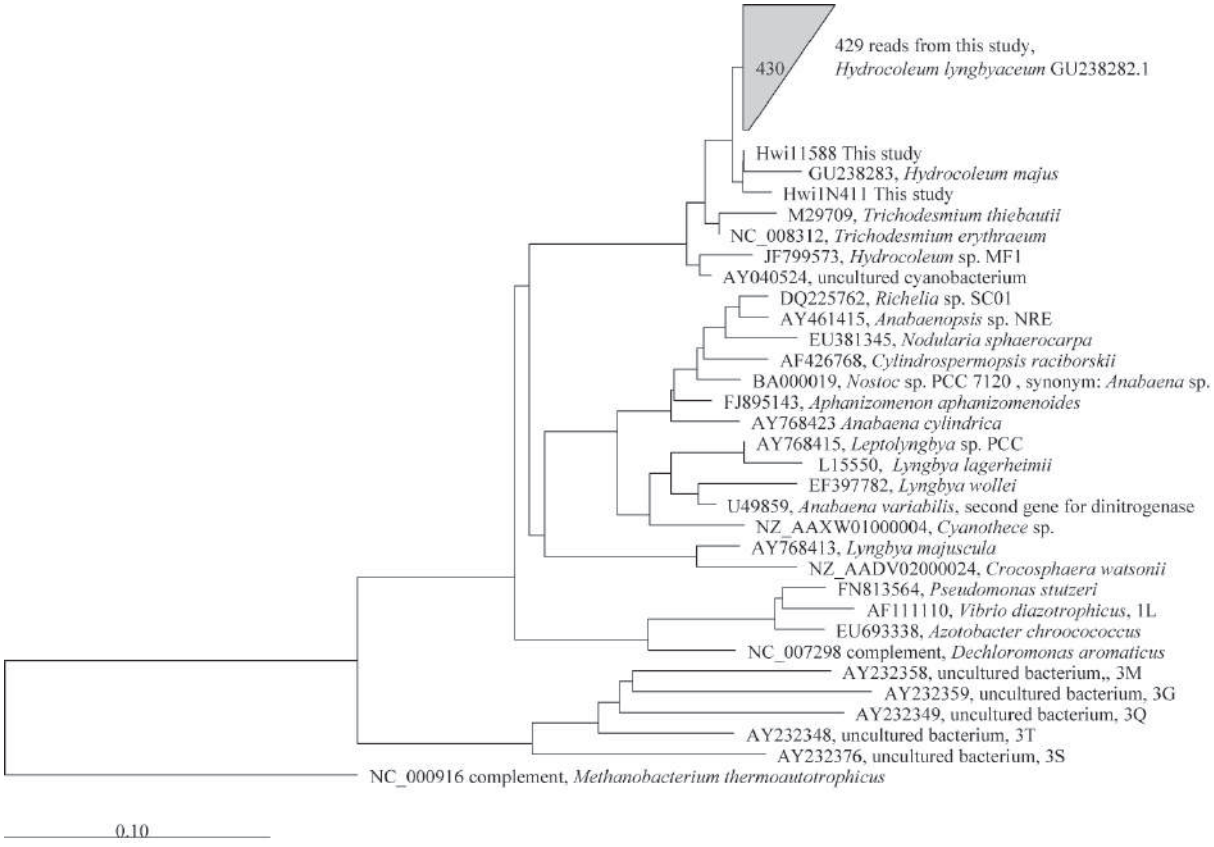
429 reads from this study,
Hydrocoleum lyngbyaceum GU238282.1
430
Hwi11588 This study
GU238283, Hydrocoleum majus
Hwi1N411 This study
M29709, Trichodesmium thiebautii
NC_008312, Trichodesmium erythraeum
JF799573, Hydrocoleum sp. MF1
AY040524, uncultured cyanobacterium
DQ225762, Richelia sp. SC01
AY461415, Anabaenopsis sp. NRE
EU381345, Nodularia sphaerocarpa
AF426768, Cylindrospermopsis raciborskii
BA000019, Nostoc sp. PCC 7120 , synonym: Anabaena sp.
FJ895143, Aphanizomenon aphanizomenoides
AY768423 Anabaena cylindrica
AY768415, Leptolyngbya sp. PCC
L15550, Lyngbya lagerheimii
EF397782, Lyngbya wollei
U49859, Anabaena variabilis, second gene for dinitrogenase
NZ_AAXW01000004, Cyanothece sp.
AY768413, Lyngbya majuscula
NZ_AADV02000024, Crocosphaera watsonii
FN813564, Pseudomonas stuzeri
AF111110, Vibrio diazotrophicus, 1L
EU693338, Azotobacter chroococcoccus
NC_007298 complement, Dechloromonas aromaticus
AY232358, uncultured bacterium,, 3M
AY232359, uncultured bacterium, 3G
AY232349, uncultured bacterium, 3Q
AY232348, uncultured bacterium, 3T
AY232376, uncultured bacterium, 3S
NC_000916 complement, Methanobacterium thermoautotrophicus
0.10

Figure 3.

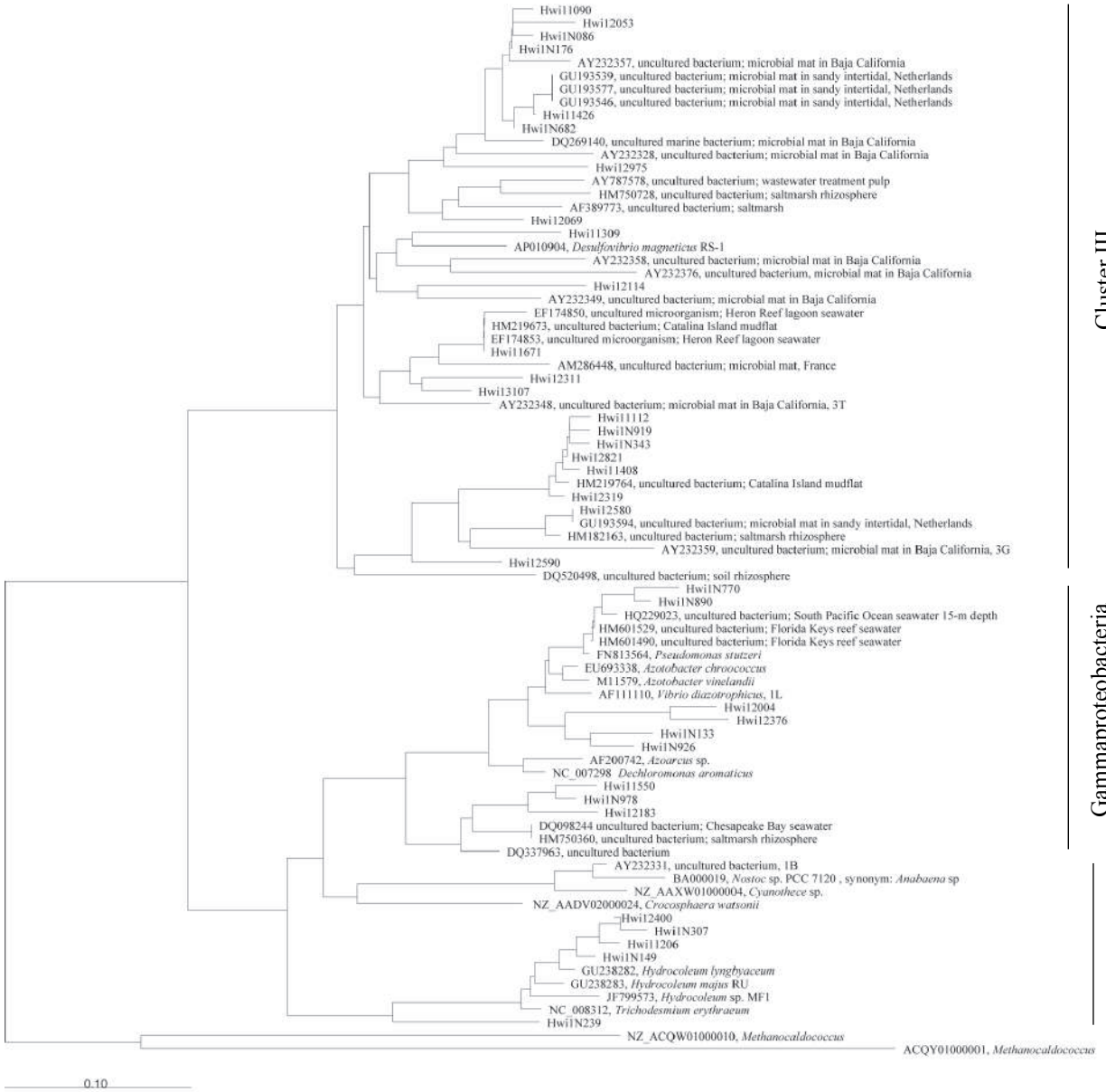
Hwi11090
Hwi12053
Hwi1N086
Hwi1N176
AY232357, uncultured bacterium; microbial mat in Baja California
GU193539, uncultured bacterium; microbial mat in sandy intertidal, Netherlands
GU193577, uncultured bacterium; microbial mat in sandy intertidal, Netherlands
GU193546, uncultured bacterium; microbial mat in sandy intertidal, Netherlands
Hwi11426
Hwi1N682
DQ269140, uncultured marine bacterium; microbial mat in Baja California
AY232328, uncultured bacterium; microbial mat in Baja California
Hwi12975
AY787578, uncultured bacterium; wastewater treatment pulp
HM750728, uncultured bacterium; saltmarsh rhizosphere
AF389773, uncultured bacterium; saltmarsh
Hwi12069
Hwi11309
AP010904, Desulfovibrio magneticus RS-1
AY232358, uncultured bacterium; microbial mat in Baja California
AY232376, uncultured bacterium, microbial mat in Baja California
Hwi12114
AY232349, uncultured bacterium; microbial mat in Baja California
EF174850, uncultured microorganism; Heron Reef lagoon seawater
HM219673, uncultured bacterium; Catalina Island mudflat
EF174853, uncultured microorganism; Heron Reef lagoon seawater
Hwi11671
AM286448, uncultured bacterium; microbial mat, France
Hwi12311
Hwi13107
AY232348, uncultured bacterium; microbial mat in Baja California, 3T
Hwi11112
Hwi1N919
Hwi1N343
Hwi12821
Hwi11408
HM219764, uncultured bacterium; Catalina Island mudflat
Hwi12319
Hwi12580
GU193594, uncultured bacterium; microbial mat in sandy intertidal, Netherlands
HM182163, uncultured bacterium; saltmarsh rhizosphere
AY232359, uncultured bacterium; microbial mat in Baja California, 3G
Hwi12590
DQ520498, uncultured bacterium; soil rhizosphere
Hwi1N770
Hwi1N890
HQ229023, uncultured bacterium; South Pacific Ocean seawater 15-m depth
HM601529, uncultured bacterium; Florida Keys reef seawater
HM601490, uncultured bacterium; Florida Keys reef seawater
FN813564, Pseudomonas stutzeri
EU693338, Azotobacter chroococcus
M11579, Azotobacter vinelandii
AF111110, Vibrio diazotrophicus, 1L
Hwi12004
Hwi12376
Hwi1N133
Hwi1N926
AF200742, Azoarcus sp.
NC_007298 Dechloromonas aromatica
Hwi11550
Hwi1N978
Hwi12183
DQ098244 uncultured bacterium; Chesapeake Bay seawater
HM750360, uncultured bacterium; saltmarsh rhizosphere
DQ337963, uncultured bacterium
AY232331, uncultured bacterium, 1B
BA000019, Nostoc sp. PCC 7120 , synonym: Anabaena sp
NZ_AAXW01000004, Cyanothece sp.
NZ_AADV02000024, Crocosphaera watsonii
Hwi12400
Hwi1N307
Hwi11206
Hwi1N149
GU238282, Hydrocoleum lyngbyaceum
GU238283, Hydrocoleum majus RU
JF799573, Hydrocoleum sp. MF1
NC_008312, Trichodesmium erythraeum
Hwi1N239
NZ_ACQW01000010, Methanocaldococcus
ACQY01000001, Methanocaldococcus
0.10
Cluster III
Gammaproteobacteria
Cyanobacteria

Figure 4.

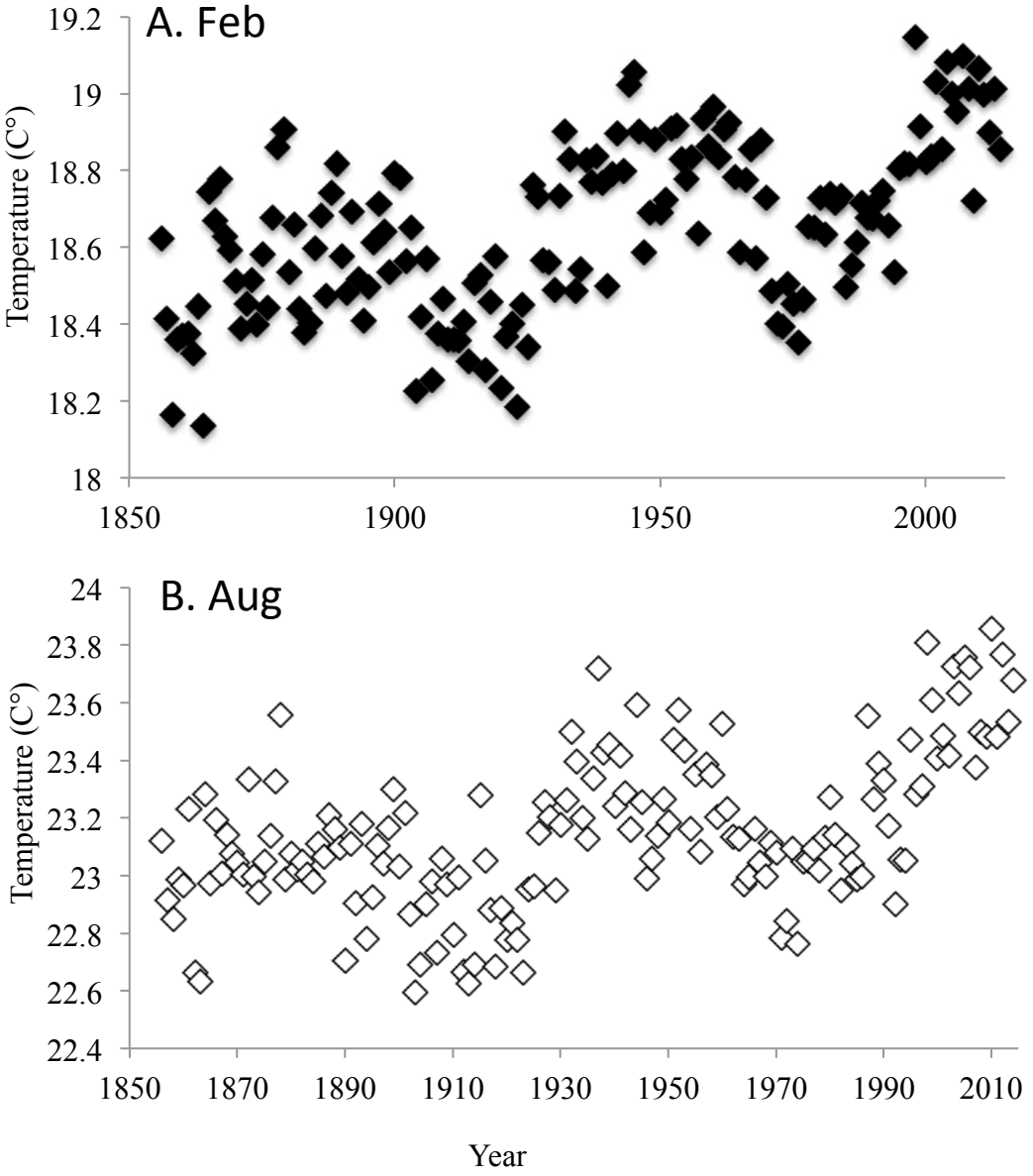
A. Feb
Temperature (C°)
19.2
19
18.8
18.6
18.4
18.2
18
1850
1900
1950
2000
B. Aug
Temperature (C°)
24
23.8
23.6
23.4
23.2
23
22.8
22.6
22.4
1850
1870
1890
1910
1930
1950
1970
1990
2010
Year

Figure 5.