

Genomic insights into the physiology and ecology of the marine filamentous cyanobacterium *Lyngbya majuscula*

Adam C. Jones^{a,1}, Emily A. Monroe^{a,1}, Sheila Podell^b, Wolfgang R. Hess^c, Sven Klages^d, Eduardo Esquenazi^a, Sherry Niessen^e, Heather Hoover^e, Michael Rothmann^f, Roger S. Lasken^g, John R. Yates III^e, Richard Reinhardt^h, Michael Kube^d, Michael D. Burkart^f, Eric E. Allen^{b,i}, Pieter C. Dorrestein^{f,j}, William H. Gerwick^{a,j}, and Lena Gerwick^{a,2}

^aCenter for Marine Biotechnology and Biomedicine, ^bMarine Biology Research Division, Scripps Institution of Oceanography, ^fDepartment of Chemistry and Biochemistry, ^dDivision of Biological Sciences, and ^hSkaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, CA 92093; ^cInstitute of Biology III, University of Freiburg, D-79104 Freiburg, Germany; ^eMax Planck Institute for Molecular Genetics, 14195 Berlin, Germany; ^gCenter for Physiological Proteomics, The Scripps Research Institute, La Jolla, CA 92037; ⁱMicrobial and Environmental Genomics, J. Craig Venter Institute, San Diego, CA 92121; and ^jGenome Centre Cologne at MPI for Plant Breeding Research, D-50829 Cologne, Germany

Edited by Jerrold Meinwald, Cornell University, Ithaca, NY, and approved April 7, 2011 (received for review January 26, 2011)

Filamentous cyanobacteria of the genus *Lyngbya* are important contributors to coral reef ecosystems, occasionally forming dominant cover and impacting the health of many other co-occurring organisms. Moreover, they are extraordinarily rich sources of bioactive secondary metabolites, with 35% of all reported cyanobacterial natural products deriving from this single pantropical genus. However, the true natural product potential and life strategies of *Lyngbya* strains are poorly understood because of phylogenetic ambiguity, lack of genomic information, and their close associations with heterotrophic bacteria and other cyanobacteria. To gauge the natural product potential of *Lyngbya* and gain insights into potential microbial interactions, we sequenced the genome of *Lyngbya majuscula* 3L, a Caribbean strain that produces the tubulin polymerization inhibitor curacin A and the molluscicide barbamide, using a combination of Sanger and 454 sequencing approaches. Whereas ~293,000 nucleotides of the draft genome are putatively dedicated to secondary metabolism, this is far too few to encode a large suite of *Lyngbya* metabolites, suggesting *Lyngbya* metabolites are strain specific and may be useful in species delineation. Our analysis revealed a complex gene regulatory network, including a large number of sigma factors and other regulatory proteins, indicating an enhanced ability for environmental adaptation or microbial associations. Although *Lyngbya* species are reported to fix nitrogen, nitrogenase genes were not found in the genome or by PCR of genomic DNA. Subsequent growth experiments confirmed that *L. majuscula* 3L is unable to fix atmospheric nitrogen. These unanticipated life history characteristics challenge current views of the genus *Lyngbya*.

diatrophism | polyketide synthase/nonribosomal peptide synthetase | mass spectrometry | harmful algal bloom | marine biology

Among the oldest life forms on Earth, cyanobacteria are well recognized for their global ecological importance and ubiquitous distribution across virtually all ecosystems (1). In the marine realm, some species of cyanobacteria contribute significantly to nitrogen fixation and global carbon flux (2), whereas others are prevalent as benthic constituents of tropical coral reefs (3). Over the past several decades, cyanobacteria have become recognized as an extremely rich source of novel, bioactive secondary metabolites (= natural products), with ~700 different compounds having been isolated and characterized (4). These compounds have gained considerable attention due to their pharmaceutical and biotechnology potential (5), but also notoriety for their environmental toxicity and threats to humans, wildlife, and livestock (6).

Marine strains of the genus *Lyngbya* are some of the most prolific producers of natural products. Nearly 240 compounds are reported from this genus, and 76% of these are attributed to a single species, *Lyngbya majuscula* (Harvey ex Gomont), which

is found globally in shallow tropical and subtropical environments (4). *Lyngbya* bloom events pose a significant challenge to coral reefs, as *Lyngbya* can negatively impact coral larvae recruitment (7), quickly colonize available substrate, and persist in the presence of herbivores because of their chemical defenses (8). In the last 10 y, focused investigations into the biosynthesis of *L. majuscula* natural products have revealed gene clusters that encode the molecular assembly of several of these compounds, including the anticancer agent curacin A (9–11), the neurotoxic jamaicamides (12), the UV-sunscreen pigment scytonemin (13), and the lyngbyatoxins (14), dermatotoxic agents responsible for “swimmer’s itch.” Most of the gene clusters encode modular, mixed polyketide synthase (PKS)/nonribosomal peptide synthetase (NRPS) assembly lines, with several using highly unusual mechanisms to incorporate other functional groups into the resultant molecules (15).

Despite these advances in compound identification and biosynthesis, comparatively little is known about *Lyngbya* evolution or the full potential of specific *L. majuscula* strains to produce the natural products attributed to this species. Recent reassessment of phylogenetic diversity in the genus *Lyngbya* using the 16S rRNA gene has shown that *Lyngbya* appears to occupy three distinct clades: a halophilic/brackish/freshwater lineage, a lineage more closely related to the genus *Oscillatoria*, and a marine lineage (16). Moreover, metabolites attributed to *L. majuscula* have typically been isolated from field collections, which poses two problems: most taxonomic classifications have been based on morphological characteristics and not genetic evidence, and this cyanobacterium typically grows in close association with other microorganisms. Thus, it is possible that the total number of natural products associated with the species *L. majuscula* has been overestimated.

To determine the capacity for natural products biosynthesis in a specific strain of *L. majuscula*, we sequenced the genome of *L. majuscula* 3L, a strain that falls within the marine lineage described earlier, and has also recently been referred to as *Lyngbya sordida* 3L (16). *L. majuscula* 3L was originally isolated in Curaçao, Netherlands Antilles, and has been maintained in stable culture

Author contributions: A.C.J., E.A.M., S.P., W.R.H., E.E., S.N., H.H., M.R., W.H.G., and L.G. designed research; A.C.J., E.A.M., S.P., W.R.H., S.K., E.E., S.N., H.H., M.R., R.S.L., R.R., M.K., and L.G. performed research; P.C.D. contributed new reagents/analytic tools; A.C.J., E.A.M., S.P., W.R.H., S.K., E.E., S.N., H.H., M.R., R.S.L., J.R.Y., M.D.B., E.E.A., P.C.D., W.H.G., and L.G. analyzed data; and A.C.J., E.A.M., S.P., W.R.H., S.N., W.H.G., and L.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The entire genome shotgun project reported in this paper has been deposited in the DDBJ/EMBL/GenBank database (accession no. [AEPQ0000000](https://doi.org/10.1093/dna/1101137108)).

¹A.C.J. and E.A.M. contributed equally to this work.

²To whom correspondence should be addressed: lgerwick@ucsd.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1101137108/-DCSupplemental.

for ~15 y (17). This strain produces curacin A (18), the molluscicidal compound barbamide (19), and the lipopeptide carmabin A (20) (Fig. S1A and B). A draft genome was obtained from an integrated strategy involving Sanger sequencing of DNA from cultured filaments in combination with 454 sequencing of DNA generated via multiple displacement amplification (MDA) from single *L. majuscula* 3L cells; the latter approach was necessary to overcome the inability to create or maintain axenic cultures of *L. majuscula*. From this sequencing effort, we aimed to confirm the presence of the gene clusters encoding each of these molecules, search for other unknown (orphan) natural product biosynthetic pathways, and gain insights into the physiological ecology of *L. majuscula* in tropical environments, including possible interactions with other microorganisms and the ability of *L. majuscula* to fix atmospheric nitrogen (21, 22).

Results and Discussion

Genome Assembly and Annotation. *L. majuscula* 3L sequence reads were obtained from two independent, nonaxenic cultures, using two different DNA isolation procedures and two different sequencing technologies (Sanger and 454 approaches). The reads from both the Sanger and 454 libraries were pooled and treated as a single metagenomic dataset to identify core sequences of the *Lyngbya* genome that were common to both datasets. This strategy enabled evaluation of whether scaffolds assembled from sequences in both datasets contained constituent reads from one or both library sources to assist in separating consensus *Lyngbya* sequences from non-*Lyngbya* contaminants. Coassembly of 712,948 Sanger and 454 reads produced 6,217 scaffolds, ranging in size from 1,000 to 59,782 nucleotides, G + C content between 25 and 76%, and coverage depth from 1- to 62-fold. Classification of 16S rRNA genes in the combined assembly, taxonomic heterogeneity of the scaffolds, and details regarding the binning procedure are available in Table S1 and Fig. S2. A total of 161 scaffolds were identified as likely originating from *L. majuscula*, on the basis of a combination of 16S rRNA genes, predicted protein matches to GenBank nr sequences, inclusion of reads from both Sanger and 454 libraries, percent G + C nucleotide composition, and assembly coverage depth. Detailed properties for all *Lyngbya*-associated scaffolds are provided in Table S2. The combined scaffolds total ~8.5 Mb, a total genome size consistent with other filamentous cyanobacteria, such as *Nostoc punctiforme* (8.2 Mb) and *Trichodesmium erythraeum* (7.8 Mb). It is uncertain whether this draft assembly represents the entire *L. majuscula* genome, but several lines of evidence suggest it is nearly complete. A survey of 102 housekeeping genes identified as nearly universal in bacteria (23) indicates that 101 of these are present in the *Lyngbya* draft genome. Copy numbers for these housekeeping genes correlate well with other sequenced cyanobacterial genomes, including those expected to have single copies (Dataset S1). In addition, previously known, independently sequenced *L. majuscula* genes for the curacin A (9) and barbamide (24) pathways are present and complete, despite not being used to guide any aspect of the assembly. The *L. majuscula* 3L draft genome was submitted to the Joint Genome Institute Integrated Microbial Genome's (IMG) expert review for automated annotation of putative ORFs. Within the 8.5-Mb genome (44% G + C content), 56 tRNAs, 2 rRNA operons, and 7,479 protein-encoding genes were identified, with 54% of these protein-encoding genes having predicted functions. This number is higher than for *N. punctiforme* (6,086 genes) and *T. erythraeum* (4,451 genes). The largest percentage of annotated genes (based on clusters of orthologous groups categories, COGs) appears to be involved in replication, recombination, and DNA repair (9%), cell wall biogenesis (8%), and signal transduction mechanisms (7%). Despite previous reports that *L. majuscula* strains are diazotrophic, no nitrogenase genes were found in this draft genome.

Secondary Metabolism Genes in *L. majuscula* 3L Draft Genome. Despite the large number of natural products attributed to *L. majuscula*, only 126 genes (3%, 293 kb) of the *L. majuscula* 3L

draft genome are predicted to be involved in secondary metabolite biosynthesis, transport, and catabolism. The majority of these are modular NRPS- and/or PKS-related genes (44%, 199 kb). Eight biosynthetic gene clusters were identified that likely encode natural products (Fig. 1 and Fig. S1). The two most apparent clusters were those of the previously characterized natural products curacin A (HQ696500) and barbamide (HQ696501) (Fig. S1A). The sequences for both pathways were complete and consistent with the sequences previously reported (9, 24). Two separate scaffolds contain genes putatively involved in carmabin biosynthesis (Fig. S1B) on the basis of predictions of adenylation domain substrate specificity (*Materials and Methods*) from the NRPS ORFs in each partial gene cluster.

Five additional biosynthetic gene clusters were found in the *L. majuscula* 3L genome; however, they do not appear to encode natural products previously detected from this species (Fig. 1). The largest of these is an apparently intact 29-kb NRPS-dominated gene cluster on scaffold 52116 that is flanked by transposase genes on both sides (HQ696495). The adenylation domain active sites of the bimodular NRPS protein are predicted to activate and incorporate proline and arginine. Surrounding the NRPS are genes for an arginosuccinate lyase, which may provide arginine for the NRPS adenylation domain, and a GCN5-related *N*-acetyltransferase (GNAT), which may acetylate arginine similarly to what has been observed with lysine acetylation in histones (25). Although a GNAT motif was described as a component of a novel PKS chain initiation mechanism for the curacin A gene cluster (10), the gene context of this motif in the current cluster appears to be different and is thus more likely to be involved in acetylation of a basic amino acid. Additionally, two adjacent phytanoyl-CoA dioxygenase (phyH)/L-proline 4-hydroxylase genes immediately precede the NRPS gene and possibly are involved in hydroxylation or halogenation of the proline residue (Fig. 1).

A second potentially complete orphan gene cluster in the *L. majuscula* 3L genome, located on scaffold 52118, is ~20 kb in size and is flanked on the 5' side by a transposase gene (HQ696496) (Fig. 1). Three NRPS ORFs, one of which is bimodular, were predicted to encode isoleucine, lysine, tyrosine, and proline. The cluster also contains a sulfotransferase, suggesting the amino acid chain could be sulfated. An epimerase domain is present in the module incorporating lysine, and thus, as in almost all known cyanobacterial metabolites containing this basic amino acid, it is likely of D configuration (4). Another separate, mixed NRPS/PKS ORF is located 13 kb downstream of this latter NRPS cluster on the same scaffold and appears to have an adenylation domain specific for either phenylalanine or tyrosine (HQ696497) (Fig. 1). From catalytic activities predicted in the PKS portion, the amino acid is likely extended with acetate, the intermediate ketone reduced to an alcohol, and then released from the enzyme by a thioesterase. The majority of the genes surrounding this stand-alone NRPS/PKS gene appear to be involved in primary metabolism, and it is unclear whether they are involved in modifications of the NRPS/PKS product.

The remaining orphan clusters are on scaffolds 52120 and 52117 (Fig. 1). Scaffold 52120 has bimodular and single module NRPS genes that have predicted adenylation specificities for α -amino adipic acid, glutamine, and proline, respectively (HQ696498). A predicted thioesterase is present at the terminus of the second NRPS gene. These are flanked by genes encoding hypothetical proteins and proteins predicted to be involved in cytochrome *c* biosynthesis. The single NRPS ORF on scaffold 52117 encodes two modules (proline and threonine adenylation specificity), and as with the NRPS/PKS on scaffold 52118, the surrounding genes appear to be related to primary metabolism (HQ696499).

To determine whether any of the above predicted "cryptic metabolites" were expressed in cultures of *L. majuscula* 3L, we first profiled water soluble and organic extracts by LC/MS (Fig. S3), matrix-assisted laser desorption/ionization (MALDI)/MS, and Fourier transform (FT)/MS. Curacin A and carmabin were readily

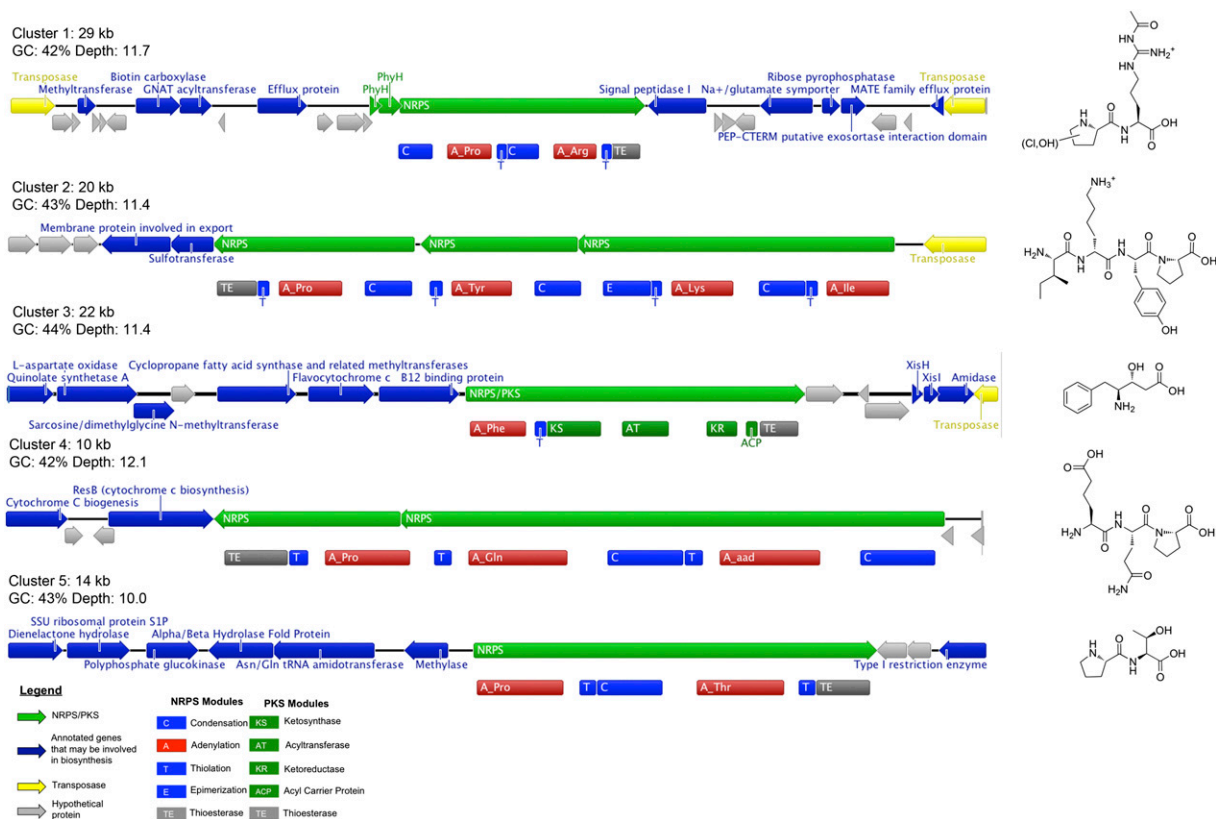


Fig. 1. Orphan gene clusters in the *L. majuscula* 3L genome with predicted product structures for each pathway. Figure constructed using Geneious 5.1 (41).

observed using all three techniques, and barbamide was detected using FT/MS. We did not detect any mass/charge values ascribable to the unknown metabolites predicted above. Using *L. majuscula* 3L soluble protein extracted from cultured biomass, we also performed a proteomic analysis to determine relative expression levels of secondary metabolite biosynthetic proteins under normal culture conditions. Multidimensional protein identification analysis (MudPIT; ref. 26) yielded spectral counts from at least two of four technical replicates for 1,043 proteins (Dataset S2), which represented ~14% of the encoded proteins annotated in the *L. majuscula* 3L genome. The most readily detected proteins using MudPIT were pigment-associated proteins, including phycocyanin subunits and phycobilisome proteins (~2,000 spectral counts per protein, Table S3A). Spectral counts for nearly all of the proteins in the curacin A and barbamide pathways were quantified, as were several proteins predicted to be involved in carminin biosynthesis (Table S3B). Only one probable secondary metabolite protein from an orphan pathway was detected in at least two technical replicates (the free-standing proline/threonine NRPS from scaffold 52,117). None of the remaining orphan pathway proteins discovered during the genome annotation process were expressed to a measurable level by MudPIT analyses. Together with the absence of any of the predicted compounds in both the water soluble and organic extracts from *L. majuscula* 3L, these data suggest that these orphan gene clusters are either expressed at low, nearly undetectable levels or not expressed at all under normal culture conditions.

Complex Regulatory Gene Network of *L. majuscula* 3L. In light of the considerable number of natural products attributed to various strains of *L. majuscula* (nearly 200 reported metabolites), it was unexpected that *L. majuscula* 3L dedicates only 3% of its genome to secondary metabolism, which is significantly lower than that observed in marine actinobacteria such as *Salinispora* (9.9%; ref. 27), and that only three NRPS/PKS-type biosynthetic pathways

larger than 30 kb were present. A larger component of the *L. majuscula* 3L genome is devoted to regulatory genes involved in transcription and signal transduction. Marine *Lyngbya* strains grow in shallow tropical areas with frequent exposure to diverse environmental stress factors such as desiccation during low tide or exposure to high fluxes of UV light. As noted previously, *Lyngbya* can usually be found living in close association with other cyanobacteria and heterotrophic bacteria. Even when growing separately from macroscopic assemblages, *L. majuscula* filaments retain a large number of associated bacterial cells on their polysaccharide sheath that are visible using DAPI staining (28). Therefore, a more careful evaluation of the *L. majuscula* 3L transcription and transduction genes was performed to evaluate the capacity of this organism for environmental adaptation and microbial communication.

L. majuscula 3L contains an unusual assortment of regulatory genes compared with other cyanobacteria. Comparison of the 15 sigma factor genes annotated in *L. majuscula* 3L against the nine well-characterized type I, II, and III sigma factors of *Synechocystis* sp. PCC 6803 (29) revealed that *L. majuscula* has precisely 1 matching sigma factor for each of the five type I and II σ^{70} factors SigA–SigE. In addition, it possesses another five factors belonging to the type III class (Fig. 2). Of this latter group, two most closely resemble SigF of *Synechocystis* sp. PCC 6803, whereas the remaining three are distinct from all other known type III factors. However, the most striking observation is the presence of 5 additional sigma factors, which have no close homolog in any of the previously sequenced model cyanobacteria *Synechocystis* sp. PCC 6803, *Anabaena* sp. PCC 7120, or *Synechococcus* sp. PCC 7942. These sigma factors are between 257 and 563 residues in length and have an unusual domain structure. A domain with pronounced similarity to σ^{24} -type factors of the extracytoplasmic function (ECF) subfamily is located in the N-terminal half of the proteins. This is intriguing because the large and diverse group of ECF sigma factors plays a key role in

adaptation to environmental conditions (30). The only other two related proteins occur as single-copy genes in the marine filamentous cyanobacteria *Trichodesmium* and *Lyngbya* sp. PCC 8106 (Fig. 2) and are annotated as SigW and Sig24 ECF-type sigma factors. However, the fact that *L. majuscula* 3L possesses five such factors suggests that a multitude of regulatory mechanisms could exist in this organism for potential interaction with the marine environment or associated microorganisms.

Moreover, the numbers and diversity of sigma factors that are global regulators of gene expression in bacteria appear higher in *L. majuscula* 3L than in most other cyanobacteria [i.e., *Anabaena* PCC 7120 has 11 sigma factors, whereas *Anabaena variabilis*, American Type Culture Collection (ATCC) 29413, and *T. erythraeum* each have 7]. In the longest ORF (HQ692799), this domain is preceded by and partially overlaps an SpvB domain (closest homolog: *Salmonella* virulence plasmid 65 kDa B protein, pfam03534). The C-terminal halves of these proteins have no close homologs in the National Center for Biotechnology Information (NCBI) or pfam databases, and they contain two 48-residue-long repeats. The similarity among four of these five proteins in their C-terminal component suggests the presence of a novel protein domain that is presently uncharacterized. ECF sigma factors are frequently cotranscribed with one or more downstream negative regulators, which function as antisigma factors that bind and inhibit the cognate sigma factor (30). The *L. majuscula* 3L ECF-type sigma factors appear to belong to a class of sigma factors in which a regulatory domain has been fused to the protein. The recently identified sigma factor PhyR in *Methylobacterium extorquens* provides a possible paradigm for such a possibility (31). In PhyR, an amino terminal ECF sigma factor-like domain is fused to a carboxyterminal receiver domain of a response regulator, suggesting PhyR can respond by sensing changes in the environment directly. A number of predicted short microRNA sequences were also found throughout the draft genome, including a cluster of mir-569 microRNA

genes, suggesting that *L. majuscula* 3L use some level of post-transcriptional regulation. Although regulation of secondary metabolism in filamentous cyanobacteria has not been extensively evaluated, we did find that homologs of two proteins to those possibly involved in jamaicamide biosynthetic regulation (32) were expressed to detectable levels according to MudPIT analysis. Whether these latter proteins are involved in the regulation of curacin A or other secondary metabolite pathways in *L. majuscula* 3L remains to be determined.

Absence of Nitrogen Fixation in *L. majuscula* 3L. Perhaps the most unexpected finding in the *L. majuscula* 3L genome analysis was the lack of any genes involved in nitrogen fixation. Nitrogen availability is thought to be a major factor regulating primary production in shallow marine environments, and fixation of atmospheric nitrogen (N_2) by some prokaryotes, including cyanobacteria, is a critical source of bioavailable nitrogen for marine ecosystems worldwide (2). Several genera of cyanobacteria have been shown to fix nitrogen, including *Lyngbya* species (21). *L. majuscula* nitrogen fixation has been detected previously by acetylene reduction assays (21, 22), and a dinitrogenase reductase (*nifH*) has been characterized from *L. majuscula* collected near Zanzibar in the Indian Ocean (22).

To independently investigate the capacity of *L. majuscula* 3L to fix nitrogen, the presence of nitrogenase genes was evaluated by PCR approaches as well as in several growth experiments performed in the absence of nitrate in the culture media. Using primers previously published to amplify *nifH* from *L. majuscula* (22), we successfully amplified a PCR product from genomic DNA isolated from *Oscillatoria nigro-viridis* 3LOSC, a cyanobacterial strain found growing in association with *L. majuscula* 3L in the field, but failed to amplify a product from *L. majuscula* 3L genomic DNA (Fig. S4). Additional experiments were performed to determine whether *L. majuscula* 3L could grow and survive in the absence of a fixed nitrogen source. Single filaments grown in nitrate-free (98%) media were comparable in length to filaments grown in normal SW BG-11 after 1 wk of growth; however, cell morphology and pigmentation were significantly altered in the nitrate-free samples. The length of cells grown in the nitrate-free media visibly increased and the filaments changed from dark red to light green and became colorless upon extended culture. Similar phenotypes were observed when this experiment was repeated using larger-scale 50-mL batch cultures (Fig. 3A). To assess nitrogen accumulation by *L. majuscula* 3L in nitrate-free media, soluble protein was isolated from 50-mL nitrate-free batch cultures after 1 wk of growth and compared with control cultures grown in control media for the same duration. During the course of the two independent experiments performed, control cultures significantly increased in protein content ($P = 0.0077$), whereas the nitrate-free cultures showed no increase in protein content ($P = 0.358$), indicating that *L. majuscula* 3L was unable to actively assimilate nitrogen from atmospheric dinitrogen (Fig. 3B).

We also explored the ability of *L. majuscula* 3L to assimilate atmospheric nitrogen through ^{15}N isotope feeding experiments. *L. majuscula* 3L filaments were grown in media containing ^{15}N -labeled sodium nitrate for ~ 21 d until nitrogen-containing compounds were fully labeled with this heavy isotope, as assessed by MALDI/TOF mass spectrometry of the metabolome (*Materials and Methods*). The fully ^{15}N -labeled filaments were then grown in nitrate-free media for 10 d, and the incorporation of the prevailing natural ^{14}N isotope from atmospheric N_2 into nitrogen-containing compounds was evaluated using MALDI/MS. The shift from ^{15}N to ^{14}N was calculated for pheophytin *a*, a chlorophyll breakdown product that readily ionizes during MALDI analysis of *Lyngbya* filaments (Fig. 3C, ref. 33). In the absence of a nitrogen source, 19% ($\pm 0.98\%$) of the pheophytin *a* shifted to lighter mass by incorporation of ^{14}N . Controls remaining in ^{15}N -labeled nitrate media showed a shift of 4% ($\pm 1.35\%$) to lighter mass, and controls grown in regular ^{14}N nitrate media shifted to lighter mass by 99% ($\pm 6.65\%$).

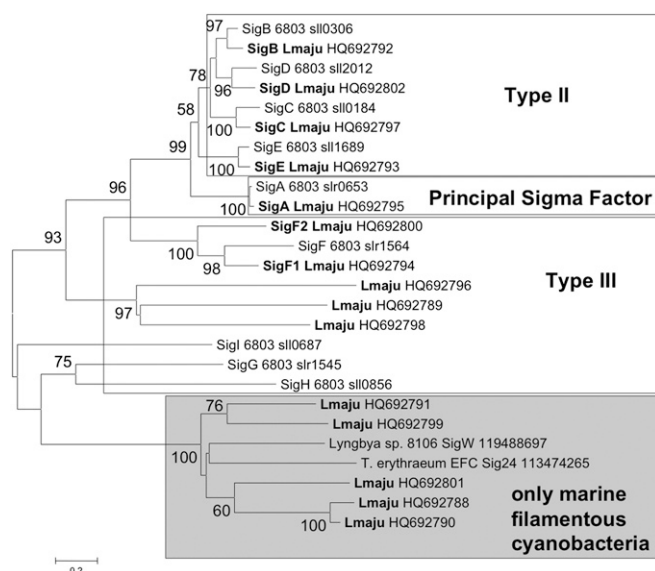


Fig. 2. Phylogenetic relationships among *L. majuscula* sigma factors. Fifteen *L. majuscula* sigma factors were compared with the 9 sigma factors from *Synechocystis* PCC 6803. The three major groups of cyanobacterial sigma factors are annotated and gene identifications are provided. Clade 4 occurs only in marine filamentous cyanobacteria and consists of 5 factors from *L. majuscula* and single representatives from *T. erythraeum* and *Lyngbya* sp. PCC8106 (minimum evolution method). The optimal tree with the sum of branch length = 14.01963947 is shown. The percentage of replicate trees in which the associated taxa clustered together (1,000 bootstrap replicates) are shown next to branches when ≥ 60 (tree to scale; branch lengths provided in same units as inferred evolutionary distances) (42).

The 19% shift observed in the nitrate-free media may be due to other trace amounts of nitrogen in the media or may represent recycling of internal nitrogen stores that were not labeled during the incubation with ^{15}N nitrate. A recent study examining proteomic changes in the cyanobacterium *Synechocystis* sp. 6803 in response to various environmental stresses, including low nitrogen, found that in addition to switching to alternative carbon and nitrogen assimilation pathways, *Synechocystis* can access internal carbon and nitrogen stores based on up-regulation of proteins associated with cyanophycin breakdown and downstream arginine catabolism (34). To provide nitrogen and carbon to the cell, cyanophycin, a storage polymer of L-aspartic acid and L-arginine found in most cyanobacteria, is broken down into arginine and aspartic acid by cyanophycinase. Arginine and aspartic acid can

be subsequently broken down by arginine decarboxylases and agmatinase and/or arginase (34). The *L. majuscula* 3L genome contains genes for a cyanophycin synthetase (HQ692807), cyanophycinase (HQ692806), two arginine decarboxylases (HQ692803 and HQ692804), and one agmatinase (HQ692805). The second arginine decarboxylase is in close proximity to the agmatinase on scaffold 52022, supporting their suggested role in cyanophycin recycling. The presence of these genes in the *L. majuscula* 3L genome provides evidence that *L. majuscula* is capable of obtaining nitrogen from cellular storage, and this capacity to use internal nitrogen stores in low nitrogen environments could explain the 19% ^{14}N shift observed in the MALDI growth experiments. The loss of pigmentation observed under nitrate-free conditions is also consistent with the observations of down-regulation of photosystem proteins in response to environmental stresses (34). Collectively, these phenotypic and growth assessments strongly suggest that *L. majuscula* is unable to fix atmospheric nitrogen, and that under nitrate-free growth conditions, recycles nitrogen from storage proteins such as cyanophycin.

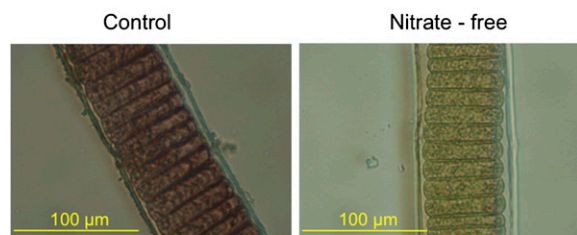
The apparent discrepancy between our investigations with *L. majuscula* 3L and past demonstrations of nitrogen fixation in *L. majuscula* may reflect strain differences or, possibly, the different criteria used in assigning the taxonomy of these cyanobacteria (21, 22). A 16S rRNA phylogenetic assessment of *L. majuscula* 3L places this strain in the marine *Lyngbya* lineage, as recently reported (16), which is phylogenetically distinct from freshwater *Lyngbya* strains. It is also conceivable that previous nitrogen fixation experiments with *L. majuscula*, wherein the organism was identified solely by morphology, may have actually investigated other morphologically similar but unrelated genera. For example, *Oscillatoria*, another cyanobacterial genus reported to fix nitrogen (35), is morphologically very similar to *Lyngbya* and can be easily misidentified without the taxonomic support provided by phylogenetics.

Conclusions

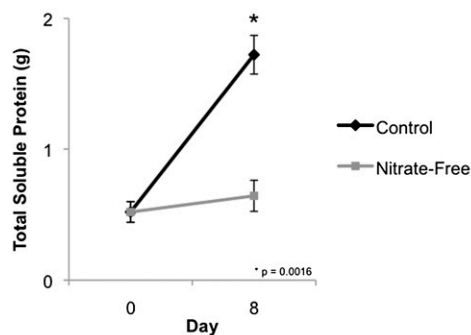
Since the first evaluation of their natural products 40 years ago, tropical filamentous marine cyanobacteria are now established as rich sources of novel bioactive molecules. We selected *L. majuscula* 3L for genome sequencing because it is a strain that has been successfully cultivated in our laboratory for 15 y and has been studied extensively for its natural products and biosynthetic pathways. The genome sequence contained intact gene clusters for curacin A and barbamide, consistent with our previous reports (9, 24), as well as genes in good agreement with carmabin biosynthesis. However, no other gene clusters above 30 kb were evident in the draft genome, and only five other PKS and/or NRPS pathways were detected. Organic extracts from *L. majuscula* 3L and expression analysis of the soluble proteome revealed that these unknown pathways are either not expressed or expressed at undetectable levels under typical culture conditions. Thus, the more than 200 metabolites reported from this species are likely due to a very large number of different chemical strains or chemotypes. Moreover, the processes of horizontal gene transfer or evolutionary pathway divergence, as suggested for the curacin A and the jamaicamide pathways from *L. majuscula* (11), or the apratoxins from strains of *Lyngbya bouillonii* (36), are likely responsible for this impressive molecular diversity.

The discovery of regulatory genes conferring an enhanced ability for microbial interactions and/or environmental adaptation and lack of traditional nitrogen fixation pathways were additional unexpected findings in the *L. majuscula* 3L genome. *Lyngbya* is almost always found growing in close association with other cyanobacteria, diverse microorganisms, and invertebrates in the field. A wide variety of heterotrophic bacteria remain on the surface of the polysaccharide sheath even after extensive purification of field isolates. The relationship between *Lyngbya* and these associated organisms remains unclear, but the possibility of complex interactions taking place among them is a fascinating focus for future research and may also explain the large number and variety of ECF sigma factors and other regulatory

A Phenotypic changes in Nitrate-Free Media



B Total Soluble Protein



C % ^{14}N Incorporation into Pheophytin a

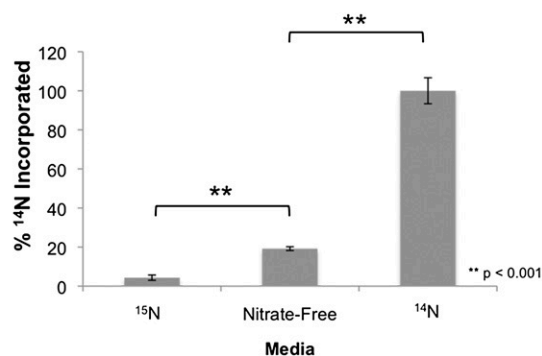


Fig. 3. Absence of nitrogen fixation in *L. majuscula* 3L. (A) Phenotypic changes of *L. majuscula* 3L when grown in nitrate-free media. Microscopic images are 400 \times . (B) Total soluble protein in *L. majuscula* 3L grown with (control, diamonds) and without (squares) nitrate for 8 d compared with day 0. Error bars represent SEM between replicate experiments ($n = 6$ per treatment). (C) ^{14}N incorporation into pheophytin a assayed by MALDI/TOF. *L. majuscula* 3L was grown in ^{15}N -labeled nitrate until nitrogen-containing molecules were fully labeled. Filaments were then grown in SW BG-11 media with ^{15}N -labeled nitrate, nitrate-free SW BG-11 media, and control ^{14}N -nitrate SW BG-11 media ($n = 8$) for 10 d and incorporation of ^{14}N into pheophytin a was measured by MALDI/TOF. Error bars are SEM.

genes described above. Our finding that nitrogen fixation does not occur in this *L. majuscula* strain is in direct contrast to previous reports, but may be another indication that finer-scale phylogenetic relationships of marine filamentous cyanobacteria need to be better defined. Among nitrogen-fixing organisms, cyanobacteria form a monophyletic group; however, within cyanobacteria, the capacity to fix nitrogen appears to be polyphyletic, suggesting multiple gene losses of nitrogen-fixation genes over time, horizontal gene transfer causing independent introduction of nitrogen-fixation genes, or a combination of both (37, 38). Recent examination of nitrogen-fixation gene evolution implies a complex history of both gene loss and horizontal gene transfer events (37, 38). The current *L. majuscula* 3L genomic data do not reveal gene loss events (via the presence of pseudogenes or regions in the genome where there was possible loss of the entire *nif* gene cluster), but as additional sequence data for diazotrophic and nondiazotrophic marine filamentous cyanobacteria become available, the evolutionary history of nitrogen fixation in this group can be better understood. Because *L. majuscula* strains previously found to fix nitrogen were identified using morphological techniques, it is difficult to determine how closely related these may be to *L. majuscula* 3L and whether the ability to fix nitrogen is more of an exception than a rule for this genus.

The first sequencing of a marine *Lyngbya* species presented here clearly accentuates the need for genomic study of additional

Lyngbya strains. Recent phylogenetic assessment has revealed a significant degree of ambiguity in *Lyngbya* taxonomy (16). At least three separate lineages have been described from different environments. Specific natural products isolated from *Lyngbya* may be a more effective means of delineating these cyanobacterial strains, as has been proposed for marine actinomycetes (39). Additional genome sequencing of other *Lyngbya* collections will be required to better understand how the traits described here compare between species, strains, and across geographic locations.

Materials and Methods

L. majuscula 3L was originally collected in 1993 near CARMABI Research Station in Curaçao, Netherlands Antilles, and live cultures have since been maintained (17) in SW BG-11 media (40). See *SI Materials and Methods* for details of fosmid library construction, single cell isolation, sequencing methods, genome assembly, binning techniques, and all other experiments. This whole genome shotgun project has been deposited at DDBJ/EMBL/GenBank under accession no. AEPQ00000000. The version described in this paper is the first version, AEPQ01000000.

ACKNOWLEDGMENTS. We thank Dr. Hyukjae Choi for assistance with LC/MS analyses and Roland Kersten for assistance with FT/MS. This work was supported by the Network of Excellence in Marine Genomics Europe, National Institutes of Health (NIH) Grants CA108874 (to W.H.G. and L.G.) and R21HG005107-02 (to E.E.A.) and California Sea Grant R/NMP-103EPD. E.A.M. is an Institutional Research and Academic Career Development Award (IRACDA) Postdoctoral Fellow supported by NIH Grant GM06852.

- Zehr JP, et al. (2001) Unicellular cyanobacteria fix N₂ in the subtropical North Pacific Ocean. *Nature* 412:635–638.
- Berman-Frank I, et al. (2001) Segregation of nitrogen fixation and oxygenic photosynthesis in the marine cyanobacterium *Trichodesmium*. *Science* 294:1534–1537.
- Thacker R, Ginsburg D, Paul VJ (2001) Effects of herbivore exclusion and nutrient enrichment on coral reef macroalgae and cyanobacteria. *Coral Reefs* 19:318–329.
- Tidgewell K, Clark BR, Gerwick WH (2010) The natural products chemistry of cyanobacteria. *Comprehensive Natural Products II Chemistry and Biology*, eds Mander L, Lui H-W (Elsevier, Oxford), Vol 2, pp 141–188.
- Tan LT (2007) Bioactive natural products from marine cyanobacteria for drug discovery. *Phytochemistry* 68:954–979.
- Carmichael WW (2001) Health effects of toxin-producing cyanobacteria: The cyanobacteria. *Hum Ecol Risk Assess* 7:1393–1407.
- Kuffner IB, Paul VJ (2004) Effects of the benthic cyanobacterium *Lyngbya majuscula* on larval recruitment of the reef corals *Acropora surculosa* and *Pocillopora damicornis*. *Coral Reefs* 23:455–458.
- Paul VJ, Thacker RW, Banks K, Golubic S (2005) Benthic cyanobacterial bloom impacts the reefs of South Florida (Broward County, USA). *Coral Reefs* 24:693–697.
- Chang Z, et al. (2004) Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 67:1356–1367.
- Gu L, et al. (2007) GNAT-like strategy for polyketide chain initiation. *Science* 318:970–974.
- Gu L, et al. (2009) Metamorphic enzyme assembly in polyketide diversification. *Nature* 459:731–735.
- Edwards DJ, et al. (2004) Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*. *Chem Biol* 11:817–833.
- Sorrels CM, Proteau PJ, Gerwick WH (2009) Organization, evolution, and expression analysis of the biosynthetic gene cluster for scytonemin, a cyanobacterial UV-absorbing pigment. *Appl Environ Microbiol* 75:4861–4869.
- Edwards DJ, Gerwick WH (2004) Lyngbyatoxin biosynthesis: Sequence of biosynthetic gene cluster and identification of a novel aromatic prenyltransferase. *JACS* 126:11432–11433.
- Jones AC, et al. (2010) The unique mechanistic transformations involved in the biosynthesis of modular natural products from marine cyanobacteria. *Nat Prod Rep* 27:1048–1065.
- Engene N, Coates RC, Gerwick WH (2010) 16S rRNA gene heterogeneity in the filamentous marine cyanobacterial genus *Lyngbya*. *J Phycol* 46:591–601.
- Rossi JV, Roberts MA, Yoo H-D, Gerwick WH (1997) Pilot scale culture of the marine cyanobacterium *Lyngbya majuscula* for its pharmaceutically useful natural metabolite curacin A. *J Appl Phycol* 9:195–204.
- Gerwick WH, et al. (1994) Structure of curacin A, a novel antimitotic, antiproliferative, and brine shrimp toxic natural product from the marine cyanobacterium *Lyngbya majuscula*. *J Org Chem* 59:1243–1245.
- Orjala J, Gerwick WH (1996) Barbamide, a chlorinated metabolite with molluscicidal activity from the Caribbean cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 59:427–430.
- Hooper GJ, Orjala J, Schatzman RC, Gerwick WH (1998) Carmabins A and B, new lipopeptides from the Caribbean cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 61:529–533.
- Jones K (1990) Aerobic nitrogen fixation by *Lyngbya* sp., a marine tropical cyanobacterium. *Eur J Phycol* 25:287–289.
- Lundgren P, Bauer K, Lugomela C, Söderbäck E, Bergman B (2003) Reevaluation of the nitrogen fixation behavior in the marine non-heterocystous cyanobacterium *Lyngbya majuscula*. *J Phycol* 39:310–314.
- Puigbò P, Wolf YI, Koonin EV (2009) Search for a 'Tree of Life' in the thicket of the phylogenetic forest. *J Biol* 8:59.
- Chang Z, et al. (2002) The barbamide biosynthetic gene cluster: A novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit. *Gene* 296:235–247.
- Yang X-J, Seto E (2008) Lysine acetylation: Codified crosstalk with other posttranslational modifications. *Mol Cell* 31:449–461.
- Washburn MP, Wolters D, Yates JR, 3rd (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 19:242–247.
- Udway DW, et al. (2007) Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc Natl Acad Sci USA* 104:10376–10381.
- Simmons TL, et al. (2008) Biosynthetic origin of natural products isolated from marine microorganism-invertebrate assemblages. *Proc Natl Acad Sci USA* 105:4587–4594.
- Imamura S, Asayama M (2009) Sigma factors for cyanobacterial transcription. *Gene Regul Syst Bio* 3:65–87.
- Helmann JD (2002) The extracytoplasmic function (ECF) sigma factors. *Adv Microb Physiol* 46:47–110.
- Francez-Charlot A, et al. (2009) Sigma factor mimicry involved in regulation of general stress response. *Proc Natl Acad Sci USA* 106:3467–3472.
- Jones AC, Gerwick L, Gonzalez D, Dorrestein PC, Gerwick WH (2009) Transcriptional analysis of the jamaicamide gene cluster from the marine cyanobacterium *Lyngbya majuscula* and identification of possible regulatory proteins. *BMC Microbiol* 9:247.
- Esquenazi E, Jones AC, Byrum T, Dorrestein PC, Gerwick WH (2011) Temporal dynamics of natural products biosynthesis in marine cyanobacteria. *Proc Natl Acad Sci USA* 108:5226–5231.
- Wegener KM, et al. (2010) Global proteomics reveal an atypical strategy for carbon/nitrogen assimilation by a cyanobacterium under diverse environmental perturbations. *Mol Cell Proteomics* 9:2678–2689.
- Stal LJ, Krumbain WE (1981) Aerobic nitrogen fixation in pure cultures of a benthic marine *Oscillatoria* (cyanobacteria). *FEMS Microbiol Lett* 11:295–298.
- Tidgewell K, et al. (2010) Evolved diversification of a modular natural product pathway: Apratoxins F and G, two cytotoxic cyclic depsipeptides from a Palmyra collection of *Lyngbya bouillonii*. *ChemBioChem* 11:1458–1466.
- Swingley WD, Blankenship RE, Raymond J (2008) Integrating Markov clustering and molecular phylogenetics to reconstruct the cyanobacterial species tree from conserved protein families. *Mol Biol Evol* 25:643–654.
- Bolhuis H, Severin I, Confurius-Guns V, Wollenzien UI, Stal LJ (2010) Horizontal transfer of the nitrogen fixation gene cluster in the cyanobacterium *Microcoleus chthonoplastes*. *ISME J* 4:121–130.
- Jensen PR, Williams PG, Oh D-C, Zeigler L, Fenical W (2007) Species-specific secondary metabolite production in marine actinomycetes of the genus *Salinispora*. *Appl Environ Microbiol* 73:1146–1152.
- Castenholz RW (1988) Culturing of cyanobacteria. *Methods Enzymol* 167:68–93.
- Drummond AJ, et al. (2010) Geneious v. 5.1. Available at www.geneious.com. Accessed October 1, 2009.
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599.