

# UC San Diego

## UC San Diego Previously Published Works

### Title

Lifestyle Evolution in Cyanobacterial Symbionts of Sponges

### Permalink

<https://escholarship.org/uc/item/8hs1c4bj>

### Journal

mBio, 6(3)

### ISSN

2161-2129

### Authors

Burgsdorf, Ilia

Slaby, Beate M

Handley, Kim M

et al.

### Publication Date

2015-07-01

### DOI

10.1128/mbio.00391-15

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-ShareAlike License, available at

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Peer reviewed

# Lifestyle Evolution in Cyanobacterial Symbionts of Sponges

Ilia Burgsdorf,<sup>a</sup> Beate M. Slaby,<sup>b</sup> Kim M. Handley,<sup>c\*</sup> Markus Haber,<sup>a</sup> Jochen Blom,<sup>d</sup> Christopher W. Marshall,<sup>e</sup> Jack A. Gilbert,<sup>c,e,f,g,h</sup> Ute Hentschel,<sup>b</sup> Laura Steindler<sup>a</sup>

Department of Marine Biology, Leon H. Charney School of Marine Sciences, University of Haifa, Haifa, Israel<sup>a</sup>; Department of Botany II, Julius-von-Sachs Institute for Biosciences, University of Würzburg, Würzburg, Germany<sup>b</sup>; Department of Ecology and Evolution, The University of Chicago, Chicago, Illinois, USA<sup>c</sup>; Bioinformatics and System Biology Justus-Liebig-University, Giessen, Giessen, Germany<sup>d</sup>; Argonne National Laboratory, Institute for Genomic and Systems Biology, Argonne, Illinois, USA<sup>e</sup>; Department of Surgery, The University of Chicago, Chicago, Illinois, USA<sup>f</sup>; Marine Biological Laboratory, Woods Hole, Massachusetts, USA<sup>g</sup>; College of Environmental and Resource Sciences, Zhejiang University, Hangzhou, China<sup>h</sup>

\* Present address: Kim M. Handley, School of Biological Sciences, The University of Auckland, Auckland, New Zealand.

I.B. and B.M.S. contributed equally to this work.

**ABSTRACT** The “*Candidatus Synechococcus spongiarum*” group includes different clades of cyanobacteria with high 16S rRNA sequence identity (~99%) and is the most abundant and widespread cyanobacterial symbiont of marine sponges. The first draft genome of a “*Ca. Synechococcus spongiarum*” group member was recently published, providing evidence of genome reduction by loss of genes involved in several nonessential functions. However, “*Ca. Synechococcus spongiarum*” includes a variety of clades that may differ widely in genomic repertoire and consequently in physiology and symbiotic function. Here, we present three additional draft genomes of “*Ca. Synechococcus spongiarum*,” each from a different clade. By comparing all four symbiont genomes to those of free-living cyanobacteria, we revealed general adaptations to life inside sponges and specific adaptations of each phylotype. Symbiont genomes shared about half of their total number of coding genes. Common traits of “*Ca. Synechococcus spongiarum*” members were a high abundance of DNA modification and recombination genes and a reduction in genes involved in inorganic ion transport and metabolism, cell wall biogenesis, and signal transduction mechanisms. Moreover, these symbionts were characterized by a reduced number of antioxidant enzymes and low-weight peptides of photosystem II compared to their free-living relatives. Variability within the “*Ca. Synechococcus spongiarum*” group was mostly related to immune system features, potential for siderophore-mediated iron transport, and dependency on methionine from external sources. The common absence of genes involved in synthesis of residues, typical of the O antigen of free-living *Synechococcus* species, suggests a novel mechanism utilized by these symbionts to avoid sponge predation and phage attack.

**IMPORTANCE** While the *Synechococcus/Prochlorococcus*-type cyanobacteria are widely distributed in the world’s oceans, a subgroup has established its niche within marine sponge tissues. Recently, the first genome of sponge-associated cyanobacteria, “*Candidatus Synechococcus spongiarum*,” was described. The sequencing of three representatives of different clades within this cyanobacterial group has enabled us to investigate intraspecies diversity, as well as to give a more comprehensive understanding of the common symbiotic features that adapt “*Ca. Synechococcus spongiarum*” to its life within the sponge host.

Received 9 March 2015 Accepted 30 April 2015 Published 2 June 2015

**Citation** Burgsdorf I, Slaby BM, Handley KM, Haber M, Blom J, Marshall CW, Gilbert JA, Hentschel U, Steindler L. 2015. Lifestyle evolution in cyanobacterial symbionts of sponges. *mBio* 6(3):e00391-15. doi:10.1128/mBio.00391-15.

**Editor** Mark J Bailey, CEH-Oxford

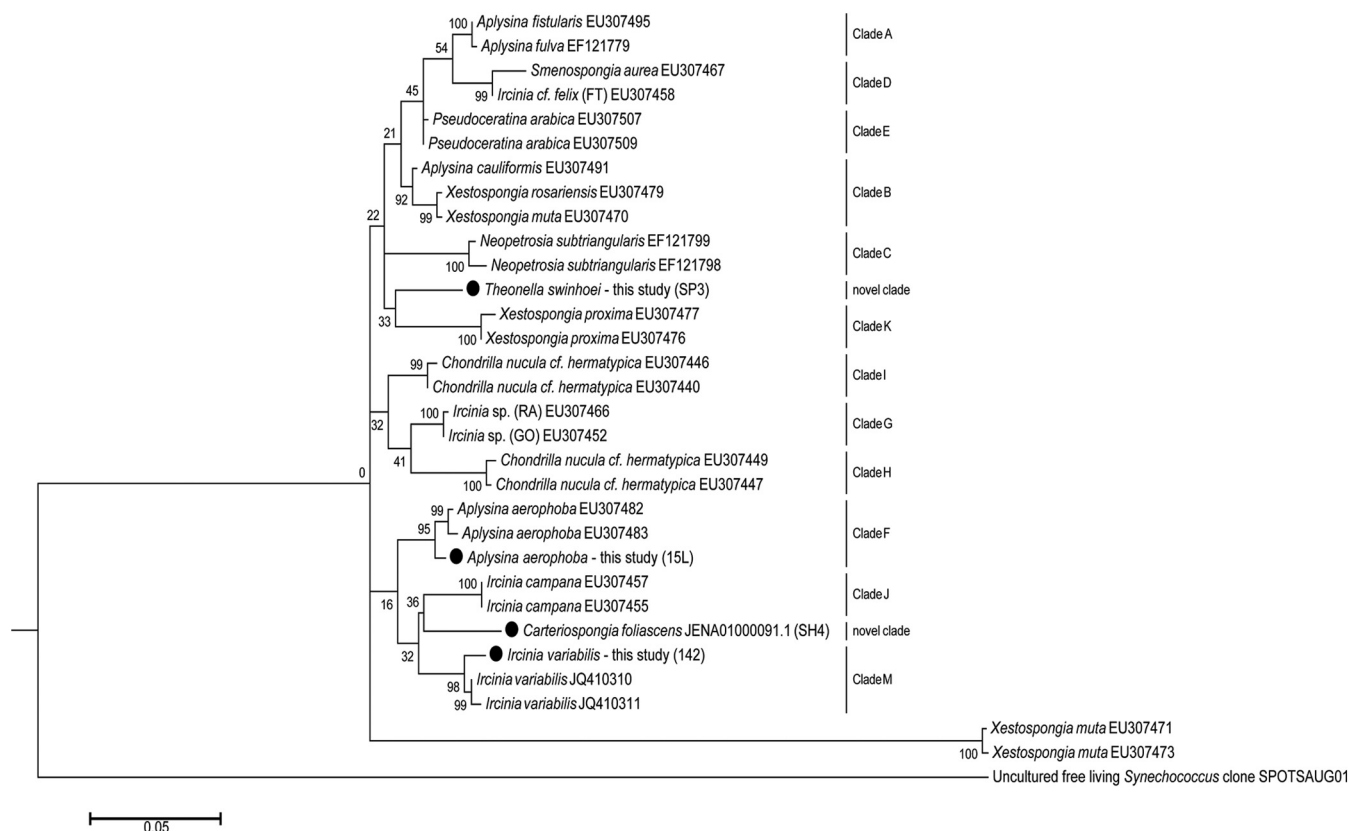
**Copyright** © 2015 Burgsdorf et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution-Noncommercial-ShareAlike 3.0 Unported license](https://creativecommons.org/licenses/by-nc-sa/4.0/), which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Address correspondence to Laura Steindler, lsteindler@univ.haifa.ac.il.

Cyanobacteria have existed as oxygenic photosynthetic bacteria on earth for at least 2.7 billion years. Over this time, they developed diverse morphologies (filamentous, unicellular, and multicellular), a plethora of physiological capacities, and a wide variety of lifestyle strategies, including symbiosis with various hosts (1–3). Cyanobacterial symbionts are polyphyletic (4) and have been reported from over one hundred different sponge species from both tropical and temperate regions (5). The major sponge-associated group of cyanobacterial symbionts is affiliated with clade VI cyanobacteria (6) and includes the nonubiquitous symbiont “*Candidatus Synechococcus feldmanni*,” found mainly in the Mediterranean and eastern Atlantic sponge *Petrosia ficiformis* (7, 8), and the widespread symbiont “*Candidatus Synechococcus spongiarum*,” which comprises at least 12 different subclades (9). The latter resides extracellularly and can be transmitted vertically to the next host generation (10–13). “*Ca. Synechococcus spongiarum*” represents an independent cyanobacterial lineage that probably became associated with sponges in the distant past and appears to be approximately equidistant from the *Synechococcus/Prochlorococcus* subclade, consisting of marine and freshwater *Synechococcus*, *Prochlorococcus*, and *Cyanobium* strains (4, 14).

Despite widespread efforts to cultivate sponge-associated marine microbes, “*Ca. Synechococcus spongiarum*” is among the vast majority of symbionts which are recalcitrant to cultivation. Next-generation sequencing and advances in bioinformatics have facilitated the recovery of genomes of uncultured sponge symbi-

ococcus spongiarum,” which comprises at least 12 different subclades (9). The latter resides extracellularly and can be transmitted vertically to the next host generation (10–13). “*Ca. Synechococcus spongiarum*” represents an independent cyanobacterial lineage that probably became associated with sponges in the distant past and appears to be approximately equidistant from the *Synechococcus/Prochlorococcus* subclade, consisting of marine and freshwater *Synechococcus*, *Prochlorococcus*, and *Cyanobium* strains (4, 14).



**FIG 1** Phylogeny of the 16S-23S ITS region (and partial 16S rRNA gene) of the sponge-associated symbiont “*Ca. Synechococcus spongiarium*.” Names on the tree are those of the host sponge species. Black circles mark sequences of genomes analyzed in this study. Maximum-likelihood criteria and distance estimates were calculated with the Kimura 2-parameter substitution model (+G+I). Bootstrap values at branch nodes derive from 1,000 replications.

onts. The first available sponge symbiont genome was that of *Cenarchaeum symbiosum*, obtained by fosmid library sequencing (15). More recently, genomes from the candidate phylum *Poribacteria* were obtained through single-cell sorting and multiple displacement amplification (16), while metagenome sequencing and contig binning resulted in the draft genomes of a sponge-associated sulfur oxidizing bacterium (17) and the first genome of “*Ca. Synechococcus spongiarium*” SH4 (14). The latter revealed a reduced genome size, an enrichment of eukaryotic-type domains, a lack of methionine precursor biosynthesis genes, and a loss of genes involved in cell wall formation and encoded low-molecular-weight peptides of photosystem II (*psb*) and antioxidant enzymes (14). How typical these traits are of all clades of “*Ca. Synechococcus spongiarium*,” inhabiting diverse sponge species and geographic locations, was previously undetermined. Earlier physiological studies reported differences in productivity and ability to assimilate and transfer carbon to the host across different sponges hosting diverse “*Ca. Synechococcus spongiarium*” clades, raising the possibility that “*Ca. Synechococcus spongiarium*”-host associations may be on different evolutionary trajectories, whereby some are obligatory and others are facultative (18, 19).

The purpose of this study was to investigate the genomic diversity among different representatives of the “*Ca. Synechococcus spongiarium*” group and to delineate the common features that are characteristic of their symbiotic existence within sponges. Three novel draft genomes, one from the Red Sea sponge *Theonella swinhoei* Gray 1868 and one each from the Mediterranean species *Ircinia*

*variabilis* Schmidt 1862 and *Aplysina aerophoba* Nardo 1833, are described here. General adaptations of this symbiont species to its sponge hosts have been revealed by comparison to free-living cyanobacteria, and intraspecies variability has been addressed through comparisons with the previously published genome of “*Ca. Synechococcus spongiarium*” SH4 (14).

## RESULTS

**Intraspecies phylogeny.** Based on BLAST homology and phylogenetic affiliation of the 16S-23S internal transcribed spacer (ITS) region (and partly the 16S rRNA gene), the three cyanobacterial genomes reported here belong to different clades of the “*Ca. Synechococcus spongiarium*” group. The maximum-likelihood phylogeny indicated that phylotype 142 belongs to clade M, supporting previous classification of this *I. variabilis* symbiont from the northwestern Mediterranean Sea (9). Phylotype 15L grouped into clade F together with other symbionts of *A. aerophoba* from previous studies (9). Phylotypes SH4 and SP3, both from the Red Sea, could not be assigned to any previously described clades with certainty and may represent novel clades (Fig. 1).

**Genome recovery.** Draft genomes of “*Ca. Synechococcus spongiarium*” were obtained for the Red Sea sponge *Theonella swinhoei* (SP3), and the Mediterranean sponges *Ircinia variabilis* (142) and *Aplysina aerophoba* (15L). The SP3 draft genome comprises 117 contigs that were cobinned based on GC content, coverage, BLAST homologies, and DNA fragment clustering within an emergent self-organizing map created with tetranucleotide fre-

**TABLE 1** General genomic information for the four “*Ca. Synechococcus spongiarum*” phylotypes 15L, SP3, 142, and SH4, and six free-living *Synechococcus* and *Cyanobium* species

Taxon <sup>a</sup>	1	2	3	4	5	6	7	8	9	10
Lifestyle (83)	Sponge symbiont	Sponge symbiont	Sponge symbiont	Sponge symbiont	NA	Euryhaline	Coastal/opportunist	Euryhaline	NA	Coastal/opportunist
Salinity	NA	NA	NA	NA	Freshwater	Halotolerant	Marine	Marine	Marine	Marine
Predicted size (Mb)	2.3	2.2	2.5	1.9	3.3	3.1	2.2	2.6	2.8	2.4
Avg GC content (%)	59.2	60.9	58.7	63.1	68.7	66.0	60.8	64.5	68.7	60.2
No. of:										
ORFs	2260	2375	2268	1792	3220	2989	2535	2679	2756	2573
Hypothetical proteins	923	1006	994	630	1182	1125	1011	1017	992	1036
SEED functions	1337	1369	1274	1162	2038	1864	1524	1662	1764	1537
SEED subsystems	264	286	237	228	326	329	313	321	331	292
COGs	1338	1332	1230	1121	2142	1931	1542	1710	1830	1578

<sup>a</sup> Taxa: 1, “*Ca. Synechococcus spongiarum*” 15L (JYFQ00000000); 2, “*Ca. Synechococcus spongiarum*” SP3 (JXQG00000000); 3, “*Ca. Synechococcus spongiarum*” 142 (JXUO00000000); 4, “*Ca. Synechococcus spongiarum*” SH4; 5, *Cyanobium gracile* PCC6307; 6, *Synechococcus* sp. strain WH5701; 7, *Synechococcus* sp. strain RCC307; 8, *Synechococcus* sp. strain RS9917; 9, *Cyanobium* sp. strain PCC7001; 10, *Synechococcus* sp. strain WH7803.

quencies. For the draft genome of 142, 327 contigs were similarly cobinned, and as only the cortex of the sponge was used for cell separation and DNA extraction, the contig coverage was clearly distinct from (far higher than) the coverages of the remaining *I. variabilis* microbial community. Similarly, only the cortex tissue of *A. aerophoba* was used to enrich for its cyanobacterial symbiont, and *Synechococcus* cells were further enriched by fluorescence-activated cell sorting (FACS) utilizing phycoerythrin and chlorophyll *a* autofluorescence as the sorting principle. This resulted in the draft genome of 15L, which consisted of 229 contigs.

The draft genomes of 142, 15L, and SP3 were found to be 91%, 95%, and 96% complete, respectively, while the previously published SH4 from the Red Sea sponge *Carteriospongia foliascens* (14) was 89% complete. This suggests that the methodological approach utilized (microbial cell enrichment by filtration and centrifugation with or without fluorescence activated cell sorting and use of the cortex with or without endosome) appeared not to be a significant determinant affecting draft genomes completeness. Predicted genome sizes ranged from 1,863,718 bp (SH4) to 2,502,041 bp (142), and GC content varied between 58.7% (142) and 63.1% (SH4) (Table 1).

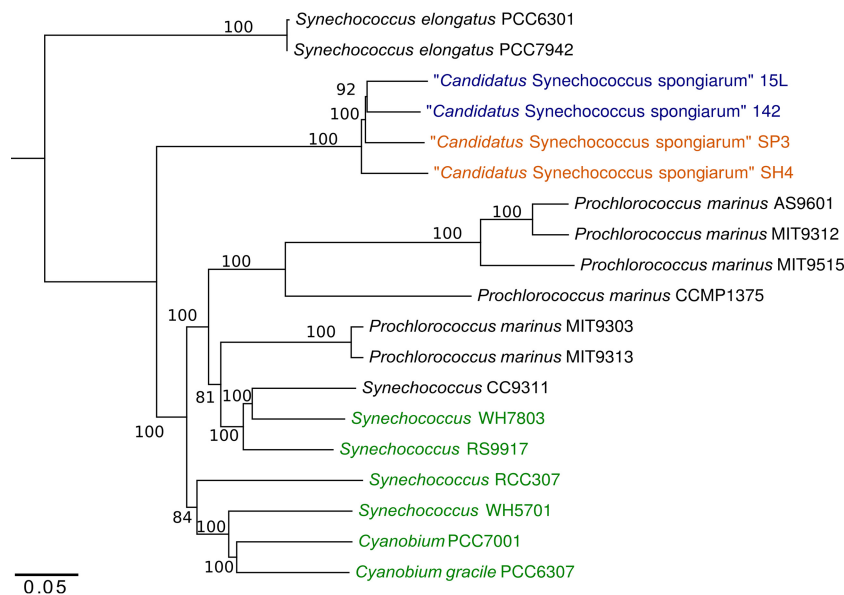
**Alignment and reordering.** With the aim of detecting lost proteins and the structural genome architecture of the four “*Ca. Synechococcus spongiarum*” draft genomes, we reordered their contigs using *Cyanobium gracile* PCC6307 as a reference. This resulted in a slight increase in detected open reading frames (ORFs) and SEED subsystems (an efficient clustering of next generation sequences). However, it should be noted the reordering is not necessarily representative of true ordering. Alignments of the four “*Ca. Synechococcus spongiarum*” genomes based on BLASTn and BLASTp analysis are shown in Fig. S1 in the supplemental material. Thereafter, the reordered genomes of SH4, 15L, and 142 were plotted against SP3, showing a high degree of within-contig gene synteny (see Fig. S2 in the supplemental material).

**Strains phylogenetically similar to “*Ca. Synechococcus spongiarum*.”** Phylogenetic analyses based either on a set of conserved marker genes (20) or on the core genome (2, 21) may provide a deeper understanding of genome evolution. For our com-

parative genome analyses, we chose six representatives of closely related free-living *Synechococcus/Cyanobium* species (shown in green in Fig. 2) and used RAST (22) to compare them to the four “*Ca. Synechococcus spongiarum*” genomes. Results from this analysis (not shown) and from a phylogenomic tree based on the core genome of the symbiotic and free-living cyanobacterial genomes (Fig. 2) support previous reports that “*Ca. Synechococcus spongiarum*” is phylogenetically equidistant from a group of marine and freshwater *Synechococcus*, *Prochlorococcus* and *Cyanobium*, termed *Synechococcus/Prochlorococcus* subclade (14).

These reference strains were further compared to the four “*Ca. Synechococcus spongiarum*” genomes based on the mean percent amino acid identity shared between orthologous genes. The mean amino acid identities within the four “*Ca. Synechococcus spongiarum*” phylotypes ranged from 91.0% to 92.1% and from 63.6% to 72.5% when “*Ca. Synechococcus spongiarum*” genomes were compared to the genomes of free-living cyanobacteria. The two highest average amino acid identities were with the marine cyanobacterium *Cyanobium* PCC7001 (72.4%) and its congeneric freshwater relative *Cyanobium gracile* PCC6307 (72.2%) (see Table S1 in the supplemental material).

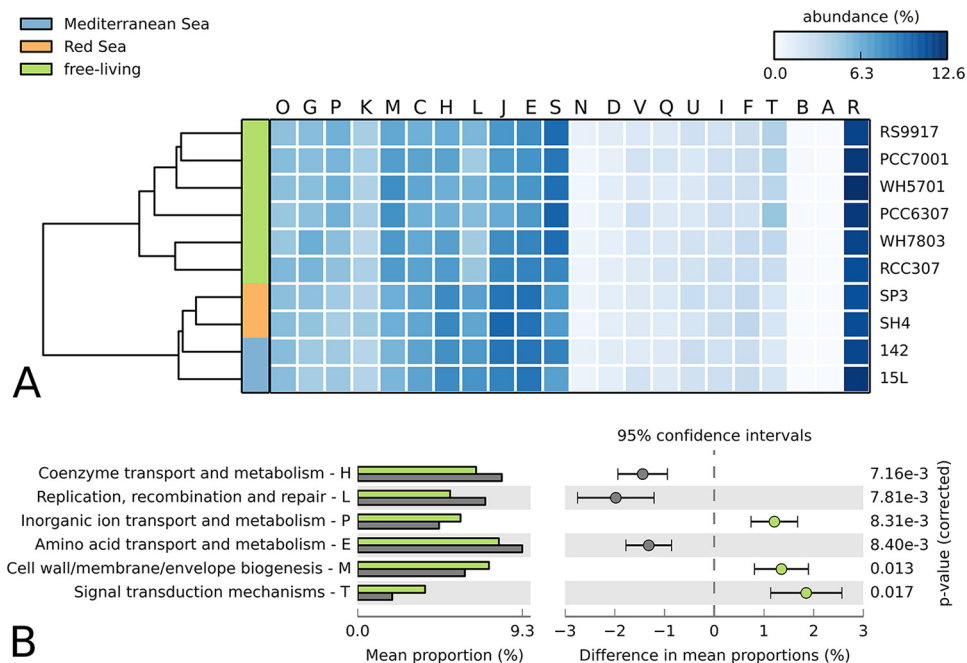
**COG functional analysis.** The genes of four “*Ca. Synechococcus spongiarum*” and six free-living cyanobacterial genomes were assigned to 1,759 clusters of orthologous groups (COGs). UP-GMA (unweighted pair group method with arithmetic average) clustering based on COG class abundances grouped the four “*Ca. Synechococcus spongiarum*” genomes together and apart from the six free-living cyanobacteria analyzed here, with the two phylotypes from Red Sea sponges (SH4 and SP3) clustering together (Fig. 3A). The “*Ca. Synechococcus spongiarum*” genomes were characterized by significantly higher proportions of COGs involved in coenzyme transport and metabolism (H), amino acid transport and metabolism (E), and replication, recombination, and repair (L) than were the free-living strains. Conversely, COGs involved in inorganic ion transport and metabolism (P), cell wall/membrane/envelope biogenesis (M), and signal transduction mechanisms (T) were more common in the free-living cyanobacteria (Fig. 3B).



**FIG 2** Concatenated phylogenetic core genome tree calculated by iterative pairwise comparison of genomes of the cyanobacteria analyzed here. Bootstrap values at branch nodes derive from 100 replications (Kimura distance matrix, neighbor joining algorithm). Names in orange and blue are “*Ca. Synechococcus spongiarum*” associated with Red Sea and Mediterranean sponges, respectively; those in green are free-living strains used for genomic comparisons.

Of the 1,759 identified COGs, 634 formed an essential functional core, present in all 10 genomes analyzed. Five hundred eighty-one COGs were unique to free-living cyanobacteria, 105 of these being present in all six genomes. One hundred twelve COGs were unique to the “*Ca. Synechococcus spongiarum*” phylotypes, 14 of these being present in all four genomes (see Table S2 in the

supplemental material). Out of these 14 COGs that were unique to the “*Ca. Synechococcus spongiarum*” core genome, four (COG2189, COG1743, COG0863, and COG0270) encoded methylases belonging to the replication, recombination, and repair class (L). Two additional COGs exclusively found in the “*Ca. Synechococcus spongiarum*” genomes were those encoding an



**FIG 3** (A) Heat map representing relative abundances of genes from COGs of different functional classes A to V. Two Mediterranean “*Ca. Synechococcus spongiarum*” genomes (blue), two Red Sea “*Ca. Synechococcus spongiarum*” genomes (orange), and six genomes of free-living cyanobacteria (green) were compared in this analysis. UPGMA clustering is presented to the left of the map. (B) COG classes with statistically significant differences between four “*Ca. Synechococcus spongiarum*” genomes (grey) and six genomes of free-living cyanobacteria (green). Error bars indicate within-group standard deviations. Presented categories passed a corrected *P* value of  $<0.05$  in Welch’s *t* test.



ankyrin repeat protein (COG0666; 4 genes annotated with this COG in each genome) and a leucine-rich repeat (LRR) protein (COG4886). COG0666 has been reported as a sponge microbe-specific signature in previous metagenomic studies (23, 24).

The “*Ca. Synechococcus spongiarum*” genomes (except that of 15L) contained leucine-rich-repeat (LRR) proteins (COG4886) in different amounts. Genome 142 harbored 13 LRR proteins (ranging between 1,577 and 4,784 bp in length), which were annotated based on the KEGG database as carboxypeptidase N regulatory subunit (K13023), surface adhesion proteins (K12549), and the extracellular matrix receptor interaction proteins (K06260). Only one LRR protein (COG4886) was found in SH4, annotated as an extracellular matrix receptor interaction protein (K06260), and this 4,463-bp-long gene consisted of seven LRR\_8 (pfam13855) and two Calx\_beta (pfam03160) domains. SP3 harbored genes for three LRR proteins (COG4886), ranging between 947 and 1,688 bp in length, which were all annotated as carboxypeptidase N regulatory subunit (K13023).

All four “*Ca. Synechococcus spongiarum*” phylotypes contained COG1629, an outer membrane receptor protein, likely used for iron transport. Meanwhile, the outer membrane receptor for ferrienterochelin and colicins (COG4771) was found only in phylotype 142. Both COG1629 and COG4771 were adjacent in genome 142 and are related to TonB-dependent receptors (SEED annotation; KEGG iron sensing pathway K02014). In addition, COG1629-like genes in SH4 and 142 showed similarity to a TonB-dependent siderophore receptor (RefSeq DELTA-BLAST e-values, 7.37e-27 and 1.79e-10, respectively). The iron-sensing pathway (K02014) was found to be absent in all six free-living cyanobacteria analyzed and is potentially absent in the *Synechococcus/Prochlorococcus* subclade. Beyond potential iron-sensing capacity, SP3 and 15L contained the largest protein conglomeration related to the K02014 pathway. Apart from the above-identified outer membrane iron receptor (COG1629), this comprised an ABC-type hemin transport system with a periplasmic component (COG4558), an ABC-type Fe<sup>3+</sup> siderophore transport system with a permease component (COG0609), and an ABC-type cobalamin/Fe<sup>3+</sup>-siderophore transport system with ATPase components (COG1120). The presence of COG1629 (K02014), COG0609 (K02015), and COG1120 (K02013) was confirmed by parallel RPSBLAST against KEGG.

Of the 105 COGs unique to the six free-living cyanobacteria, 40 belonged to the COG classes “general function prediction only” (R) and “unknown function” (S), 9 belonged to “replication, recombination and repair” (L), and 8 belonged to “cell wall/membrane/envelope biogenesis” (M). The classes “translation, ribosomal structure and biogenesis” (J) and “inorganic ion transport and metabolism” (P) consisted of 5 and 6 COGs, respectively. None of the other COG classes had more than 4 COGs unique to the free-living cyanobacteria (data not shown).

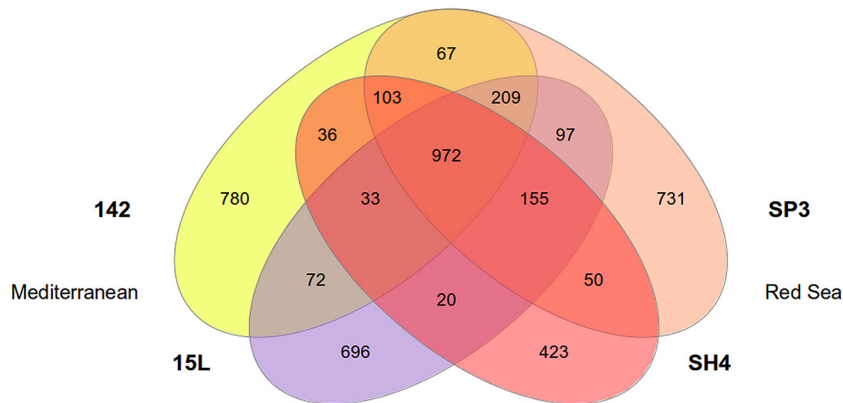
STRING, a protein-protein interaction network that predicts protein associations, was used to obtain potential linkages between different COGs present only in the free-living cyanobacterial genomes (25). One network revealed by this analysis included six COGs, five of them belonging to the class “cell wall/membrane/envelope biogenesis” (M) and one to “carbohydrate transport and metabolism” (G). These six COGs were annotated as encoding dTDP-4-dehydrorhamnose reductase (COG1091), dTDP-glucose pyrophosphorylase (COG1209), dTDP-4-dehydrorhamnose 3,5-epimerase (COG1898), dTDP-D-glucose 4,6-dehydratase

(COG1088), GDP-D-mannose dehydratase (COG1089), and mannose-6-phosphate isomerase (COG0662). Among the above-mentioned COGs, COG1089 is important for the production of GDP-D-rhamnose, while COG1091, COG1209, COG1898, and COG1088 are part of a pathway that processes D-glucose 1-phosphate to dTDP-4-dehydro-beta-L-rhamnose. The lack of these genes was also confirmed by RAST analysis (EC 4.2.1.47, EC 2.7.7.24, EC 5.1.3.13, and EC 1.1.1.133). L-Rhamnose is one of the important residues of the O antigen of lipopolysaccharides (LPS) in Gram-negative bacteria, has been previously detected in the LPS of free-living marine *Synechococcus* (26), and may be involved in host-microbe recognition and/or phage resistance.

**Symbiotic minimalism.** COG-based analysis also revealed that the four “*Ca. Synechococcus spongiarum*” genomes harbored fewer genes related to several essential functions than their free-living counterparts. Examples of reduced genes included signal transduction (COG0642), transcriptional regulation (COG0745 and COG0664), ABC-type phosphate transport (COG0226), carotenoid biosynthesis (COG3239 and COG1233), translation and posttranslational modification related proteins (COG0459, COG0229, and COG0695), carbohydrate transport and metabolism (COG1175, COG0363, and COG0366), and subunits of cytochrome *c* (COG1845, COG1622, and COG0843) (see Table S3 in the supplemental material). It is noteworthy that the organelle of *Paulinella chromatophora*, representing an auxiliary endosymbiotic acquisition (27), also showed partial reduction of a number of genes related to these essential functions (see Table S3 in the supplemental material).

**BLAST-based gene comparisons of symbionts versus free-living organisms.** The pangenome of the four “*Ca. Synechococcus spongiarum*” consisted of 3,726 genes; the core genome consisted of only 972 genes (Fig. 4). BLAST-based analysis with EDGAR (21) revealed 173 genes that were present in all four phylotypes of “*Ca. Synechococcus spongiarum*” but absent from all six free-living cyanobacteria, suggesting that they may contain potentially unique symbiotic features. Only three of these had COG annotations that were unique to “*Ca. Synechococcus spongiarum*” (COG5395, COG1651, and COG2932) (see Table S4 in the supplemental material). COG5395 (a predicted membrane protein of unknown function, belonging to superfamily DUF2306) was not reported in any *Synechococcus*, *Prochlorococcus*, *Synechocystis*, or *Cyanobium* strains. COG1651 (encoding a disulfide interchange protein) was found only in freshwater *Synechocystis* sp. strain PCC6803. COG2932 (encoding a predicted transcriptional regulator) was also found only in *Synechococcus* strain JA33, a nonmarine strain. This gene was annotated as COG2932 in 15L, SP3, and SH4, while in 142 it was annotated as COG0681 (signal peptidase I).

The 173 genes were filtered by omitting the genes whose COG annotation was not unique to “*Ca. Synechococcus spongiarum*” (i.e., COGs present in some or all of the six genomes from free-living cyanobacteria analyzed). Out of the resulting 78 genes, 75 had no COG classification (see Table S4 in the supplemental material). Of these, 49 were hypothetical genes according to SEED and did not get KEGG annotations either. The remaining 22 genes included a tetratricopeptide repeat (TPR) (K07280), phycoerythrin-associated proteins (K05380 and K05379), phyco-cyanobilin:ferredoxin oxidoreductase (K05371), allophycocyanin subunit (K02092), and nickel-dependent superoxide dismutase (EC 1.15.1.1).

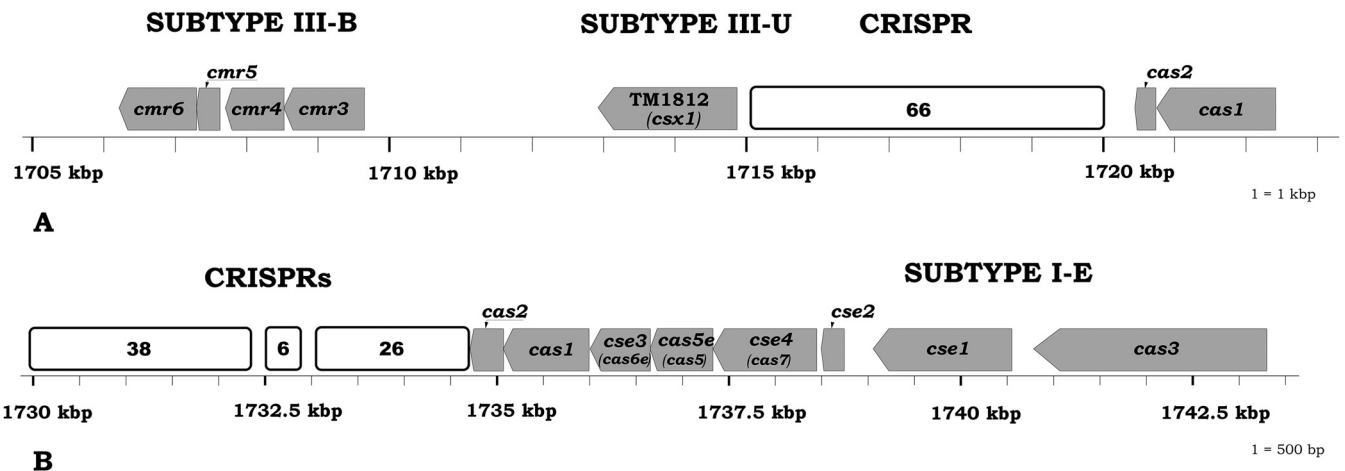


**FIG 4** Venn diagram comparing the gene inventories of four “*Ca. Synechococcus spongiarum*” genomes computed by EDGAR (21) based on reciprocal best BLAST hits of the coding sequences predicted by RAST (22). SH4 and SP3 are symbionts of Red Sea sponges, and 15L and 142 are symbionts of Mediterranean sponges.

**CRISPR-Cas systems.** Previous studies reported an almost complete lack of clustered regularly interspaced short palindromic repeats (CRISPR)-Cas systems within the *Synechococcus/Prochlorococcus* subclade (the only exception being *Synechococcus* sp. strain WH8016) (28). In contrast, we found CRISPR-associated proteins in all four “*Ca. Synechococcus spongiarum*” genomes and CRISPR regions in three of them (142, 15L, and SH4). Eight unambiguous CRISPR regions were detected in phylogroup 142, including two large CRISPR modules (see Table S5 in the supplemental material), one with a single spacer region of 66 spacers (module 1) and another with 3 spacer regions with 70 spacers (module 2) (Fig. 5). The modules were separated by a gap of more than 7.5 kb, and each included Cas1 and Cas2 proteins upstream of the CRISPR region. The CRISPR region of module 1 (4,973 bp) had a 420-bp gap of noncoding DNA (suspected leader sequence) between it and the Cas1-Cas2 operon. An operon of genes encoding CRISPR-associated proteins belonging to the Cmr group (RAMP superfamily) was found more than 3 kb downstream of this module. We attributed this operon to module 1. All

RAMP family proteins belonged to subtype III-B according to the classification by Makarova and colleagues (29). TM1812, another CRISPR-associated protein, was found 170 bp downstream of the CRISPR region. We attributed it also to module 1 and to subtype III-U. No gap was present between the Cas1-Cas2 operon and the associated CRISPR regions in module 2. Three CRISPR regions with 38, 6, and 26 spacers formed the CRISPR part of module 2 and had common lengths of 4,458 bp. All three CRISPR regions had a similar spacer sequence. The gaps between those three regions were 144 and 145 bp. Module 2 included an operon of CRISPR-associated proteins belonging to subtype I-E. A helicase (COG1200) was found upstream of the CRISPR-associated protein conglomeration in the 142 genome. A phage-related regulatory protein cII gene (COG1192) was adjacent to helicase COG1200. Furthermore, an additional *cas2* gene was found dissociated from both modules.

SH4 contained only two CRISPR regions, each including six spacers. The CRISPR-associated proteins Cse 4, 2, 1 and Cas1 formed a conglomeration. However, CRISPR regions and



**FIG 5** Schematic representation of the genomic architectures of two CRISPR-Cas of “*Ca. Synechococcus spongiarum*” 142. The number of spacers of the CRISPR regions and the closest CRISPR-Cas subtype according to Makarova and colleagues (29) are shown. The names of genes are described as they were annotated in the analysis (see Materials and Methods). The names in parentheses were added when the annotated gene names differed from the nomenclature proposed by Makarova and colleagues (29). (A) Module 1, consisting of proteins resembling subtype III-B and subtype III-U. (B) Module 2, showing proteins resembling subtype I-E.

CRISPR-associated proteins did not form a module and were found on different contigs and likely in different parts of the genome. Like in the genome of 142, additional helicases (COG1247) and phage-related regulatory protein cII (COG1192) were found upstream of the CRISPR-associated proteins conglomeration. In 15L, one CRISPR region was detected containing 49 spacers and also a conglomeration of Cas1, Cas4, Cas2, and two Cas3 proteins. The genome of SP3 appeared devoid of CRISPR regions, but a conglomeration of Cas1, Cas4, and two Cas3 proteins was detected. All six free-living cyanobacteria lacked CRISPR regions and CRISPR-associated proteins, as expected from a previous study (28).

**Functional features at the genomic level.** RAST-based analysis revealed the near completeness of key functional pathways in the four “*Ca. Synechococcus spongiarum*” genomes. These included glycolysis, the tricarboxylic acid (TCA) cycle, nitrogen metabolism, the pentose phosphate pathway, fatty acid biosynthesis, fructose and mannose metabolism, amino sugar metabolism, pyruvate metabolism, amino acid metabolism, sulfur metabolism, sucrose metabolism, and photosynthesis. The lack of two enzymes related to L-homocysteine synthesis (a methionine precursor) was reported before for SH4 (14). The missing enzymes are adenosylhomocysteinase (EC 3.3.1.1; transforming S-adenosyl-L-homocysteine to the methionine precursor L-homocysteine) and O-acetyl-L-homoserine acetate-lyase (EC 2.5.1.49; involved in synthesis of L-homocysteine from L-homoserine) (14). Reanalysis of the SH4 genome using SEED and RPSBLAST against KEGG confirmed this result and determined a lack of EC 2.5.1.49 also in SP3 and 142. Conversely, the enzyme for *de novo* biosynthesis of homocysteine from S-adenosyl-L-homocysteine was detected in the three “*Ca. Synechococcus spongiarum*” draft genomes from *A. aerophoba*, *I. variabilis*, and *T. swinhoei*. The methionine salvage pathway was only partially present in “*Ca. Synechococcus spongiarum*,” with five enzymes missing in 15L, SP3, and 142 and seven enzymes missing in SH4 (see Table S6 in the supplemental material). All the missing enzymes were present in the genomes of free-living cyanobacteria analyzed.

**Loss of low-molecular-mass peptides in photosystem II.** In general, the previously reported loss of small peptides (*psb* genes) in SH4 (14), in contrast to free-living cyanobacteria, was confirmed here for all four “*Ca. Synechococcus spongiarum*” genomes analyzed (see Table S7 in the supplemental material). Our analysis confirmed the previously reported absence of *psbY*, *psbP*, and *psbK* genes in SH4. While *psbY* was also absent in 142 and 15L, we found a hypothetical copy (supported only by KEGG annotation) in SP3. The gene *psbM*, not annotated by SEED and KEGG, was detected by BLAST analysis. Though *psbI* was reported as missing in SH4 (14), we found the gene coding for this peptide with SEED annotation, with the exception of 15L. On the other hand, *psbD* and *psbP*, both reported as present in SH4 (14), were missing in all four “*Ca. Synechococcus spongiarum*” genomes according to our analysis. In contrast, two copies of *psbD* and one copy of *psbP* were present in all six free-living cyanobacteria. SH4 lacked the genes *psbH*, *psbN*, and *psb28*, whereas one copy of each of these genes was present in SP3, 15L, and 142.

**Oxidative stress.** Between 13 and 15 different genes related to resistance to oxidative stress were found in “*Ca. Synechococcus spongiarum*,” while 23 to 29 such genes were present in the six analyzed free-living cyanobacteria. Glutathione peroxidase (EC 1.11.1.9) was the only gene of this category missing in all four “*Ca.*

*Synechococcus spongiarum*” genomes, while it was present in all six free-living cyanobacteria (see Table S8 in the supplemental material).

## DISCUSSION

“*Ca. Synechococcus spongiarum*” is likely the most widespread cyanobacterial symbiont found in marine sponges. Its association with a wide variety of host species from distant geographic locations raises the question of how conserved its genome is. Recent studies suggested that genetically distinct clades of “*Ca. Synechococcus spongiarum*” photosymbionts differ in their productivity and ability to assimilate carbon and transfer it to the host sponge (19). Such different physiology may point to a diverse genome composition among clades. Here, we analyzed four genomes that represent four different clades of “*Ca. Synechococcus spongiarum*” and are associated with four different host sponges from two geographic locations. The four genomes, while sharing more than 98.6% identity at the 16S rRNA gene level, shared only 972 protein-coding genes, which account for approximately half of the total number of coding genes per genome, suggesting high variability between clades and specific adaptations for each symbiont type. Conversely, a coastal and an off-shore strain of diazotrophic cyanobacterial symbionts (UCYN-A) of prymnesiophyte algae, with 98.7% identity at the 16S rRNA level, share a higher proportion of their genome (96.6% common genes), suggesting that they can be grouped at the same functional level (30). However, within core genomes, the average amino acid sequence identity between orthologous genes is lower between the two UCYN-A strains (86%) than among the 4 “*Ca. Synechococcus spongiarum*” clades (>91%). Unfortunately, the high number of genes of unknown function in “*Ca. Synechococcus spongiarum*” limits our ability to understand the significance of much of the genomic divergence.

The first sequenced genome of “*Ca. Synechococcus spongiarum*,” SH4 from the Red Sea sponge *C. foliascens*, was reported to lack biosynthetic genes for the methionine precursor, to have a reduced number of genes involved in the formation of the Gram-negative cell wall, and to have lost several genes encoding low-molecular-weight peptides of photosystem II (*psb*) and genes coding for antioxidant enzymes (14). Our study provided evidence that most of these traits are common to multiple clades of “*Ca. Synechococcus spongiarum*,” while it also revealed novel features that appear to be clade specific.

**Sponge-specific functional genomic signatures in the “*Ca. Synechococcus spongiarum*” group.** Previous global metagenomic analyses of sponge microbiomes provided evidence for functional genomic signatures that clearly separate them from bacterial communities in the surrounding seawater (23, 24) and were able to link some genomic traits to specific members of the heterotrophic bacterial sponge community, through binning of metagenomic contigs into five draft genomes (23, 31). Here, we determined which of the previously reported typical sponge symbiont genomic signatures are also present in the ubiquitous cyanobacterial sponge symbiont “*Ca. Synechococcus spongiarum*.” One example is the significantly higher proportion of “recombination and repair” (L) COGs identified in “*Ca. Synechococcus spongiarum*” genomes. This characteristic is considered important for the stable insertion of mobile DNA into the chromosomes with repair of flanking regions in sponge-associated microbial communities (23, 24). A high presence of transposable insertion elements has also been reported for bacterial symbionts of very



different host types, such as the intracellular symbiont *Wolbachia pipientis* wMel of *Drosophila melanogaster* (32), and may be a driving force for the evolutionary adaptation of microbial populations to specific niches (33, 34). Three of the four “*Ca. Synechococcus spongiarum*” genomes possess the transposase COG3293, previously found to be enriched in sponge microbiome metagenomes compared to planktonic bacterial metagenomes (24). Moreover, COG3293 and the site-specific DNA methylase COG0270 represent highly conserved sequences that suggest the importance of these horizontal gene transfer features.

Sponge microbial metagenomes were shown to be rich in proteins containing eukaryotic-type domains, such as ankyrin and tetratricopeptide repeats (TPR), involved in protein-protein interactions in eukaryotes, and leucine-rich-repeat (LRR) domains (23, 24). LRR proteins were previously found to be essential for virulence in the pathogen *Yersinia pestis* (35) and can activate host cell invasion by the pathogen *Listeria monocytogenes* (36). Gao and colleagues reported the presence of LRRs and ankyrin repeats in SH4, while a higher frequency of TPR domains was not detected (14). When expressed in *Escherichia coli*, sponge symbiont-derived ankyrin repeat proteins can modulate phagocytosis of amoebas. This has been used as model system, as it resembles sponge amoebocytes (37). All four “*Ca. Synechococcus spongiarum*” genomes contained four representatives of the ankyrin repeat protein gene COG0666, while the six free-living cyanobacteria contained none. The genome of a sponge symbiont sulfur-oxidizing bacterium from the sponge *Haliclona cymaeformis* likewise contains a large number of ankyrin domains (17). Ankyrin repeat domains probably represent an obligatory feature for diverse sponge bacterial symbionts, yet they do not appear to be unique to sponge symbionts, as they are also found to be abundant in other symbiotic systems, such as in the genome of the above-mentioned intracellular symbiont *W. pipientis* of *D. melanogaster* (32).

Another abundant feature previously identified in sponge microbial metagenomes is CRISPRs (24). CRISPRs and their associated proteins form adaptive immunity systems that are present in most archaea and many bacteria and act against invading genetic elements, such as viruses and plasmids (29). Evidence for the presence of the CRISPR-Cas system has been found in the majority of sequenced cyanobacterial genomes, with the exception of the *Synechococcus/Prochlorococcus* subclade (28). The absence of CRISPR-Cas systems in this subclade has been attributed to either the high genetic load of this phage resistance system—possibly too high for the small-genome *Synechococcus/Prochlorococcus* subclade—or the possibility that high viral diversity may outrun the CRISPR-Cas immune system, as suggested by a mathematical model (38), although the model currently lacks empirical proof. Conversely, in this study, the presence of CRISPR-Cas systems in the small-genome-sized and highly phage-exposed cyanobacterial sponge symbiont 142 suggests that the absence of CRISPR systems in the phylogenetically related free-living *Synechococcus/Prochlorococcus* subclade has an alternative explanation. The CRISPR-based immune system may have been lost from the *Prochlorococcus/Synechococcus* ancestor after the divergence from “*Ca. Synechococcus spongiarum*” (and thus the presence in “*Ca. Synechococcus spongiarum*” resembles the ancestral state) or acquired by “*Ca. Synechococcus spongiarum*” by horizontal gene transfer, possibly from other sponge-associated bacteria. Prevalent CRISPR-Cas systems in sponge

symbionts may indicate a high selective pressure for acquiring phage resistance inside sponges. Sponge-associated bacteria, given the high water pumping rates, are likely exposed to many more viral particles than their free-living counterparts, with estimated exposure rates of approximately 1,000 viral particles per bacterial cell per day (23), which could sustain retention of the CRISPR-Cas systems in these symbionts.

**Common genomic features within the “*Ca. Synechococcus spongiarum*” group.** Streamlining of genomes in symbionts (39, 40) may eventually lead to complete dependence on the host and to the evolution of organelles, as in the case of mitochondria and chloroplasts. The pattern of gene reduction involved in essential functions in “*Ca. Synechococcus spongiarum*” resembles the one recently described for the plastid of the amoeba *P. chromatophora*. For example, cytochrome *c* oxidase, carotenoid biosynthesis, and regulators involved in signal transduction were reduced both in the four “*Ca. Synechococcus spongiarum*” genomes and in the plastid *P. chromatophora* (27) (see Table S3 in the supplemental material). Although it is difficult to conclude that genes are missing in unclosed genomes, the similar trend among the four different “*Ca. Synechococcus spongiarum*” clades supports this notion.

The loss of several *psb* genes in SH4 has been interpreted as indicating a less stable PSII complex than in free-living strains, possibly representing a low-light-adapted photosynthetic system (14). We expanded this finding to three additional “*Ca. Synechococcus spongiarum*” clades and found that *psbD* and *psbP* were absent in all four “*Ca. Synechococcus spongiarum*” genomes (see Table S7 in the supplemental material). The *psbP* gene may optimize the water-splitting reaction; its deletion in “*Ca. Synechococcus spongiarum*” could lead to a less efficient photosynthetic system and a decreased competitive potential (41). *I. variabilis* was reported to harbor more than one species of cyanobacterial symbionts, and thus, competition between different cyanobacterial species may exist within a single sponge; however, the different cyanobacterial species were shown to be spatially separated, with “*Ca. Synechococcus spongiarum*” found only in the cortex and the other two cyanobacterial symbionts (“*Ca. Synechococcus feldmanni*” and “*Ca. Aphanocapsa raspaigellae*”) detected only in the sponge matrix (42). *psbY*, missing in three different members of “*Ca. Synechococcus spongiarum*,” is nonessential for oxygenic photosynthesis in *Synechocystis* sp. PCC6803 (43). The lack of nonessential photosynthetic genes may be explained by a tradeoff where decreased competitive potential is balanced by a reduced cost of genome replication owing to genome size reduction in symbiotic cyanobacteria (2, 40).

Reactive oxygen species (ROS) are a lateral product of aerobic metabolism and can cause oxidative damage to photosynthetic organisms, such as cyanobacteria, whose antioxidant enzymes facilitate oxidative stress resistance (44). SH4 was characterized by a loss of several antioxidant enzymes (14), a finding that also was confirmed in our three additional “*Ca. Synechococcus spongiarum*” genomes. It is possible that the reduced light radiation reaching photosymbionts inside the sponge tissue decreases the amount of ROS. Furthermore, the heterotrophic microbial community within the sponge, which is present in close spatial proximity to the photosynthetic symbionts, may respire part of the oxygen concomitantly while it is produced by photosynthesis, thus reducing the accumulation of ROS in the tissue.

In addition to the previously reported loss of genes involved in cell wall formation in SH4 (14), our analysis revealed a remarkable

novel aspect: the common loss of genes involved in the production of dTDP-L-rhamnose. The latter is a residue of the O antigen of LPS that has previously been detected in free-living marine *Synechococcus* (26). Different O antigens contribute to the variation of the Gram-negative bacterial cell wall. The correct structures of LPS and its O antigen are known to be essential for the establishment of disease (pathogens) or beneficial outcomes (symbionts) in host-microbe interactions (reviewed in reference 45). With planktonic cyanobacteria being part of the typical sponge diet (46), we hypothesize that O antigens, including dTDP-L-rhamnose and GDP-D-rhamnose, may be utilized as “food receptors” by the sponge. Early studies of the 1970s had already hypothesized potential mechanisms by which sponges could differentiate between symbiont and food bacteria. It was suggested that either symbionts may be specifically recognized as such (the recently discovered ankyrin proteins being a possible mechanism) (37) or they may not be recognized at all by the phagocytic sponge cells, thanks to some kind of masking coatings (47). *In situ* feeding experiments with bacterial isolates from sponges (potential symbionts) versus isolates from seawater (free-living bacteria) coupled with electron radioautography supported the masking hypothesis. It was speculated that symbiont bacteria were not recognized as consumable bacteria thanks to a masking mechanism where chemical compounds surrounding the bacteria would act as protective capsules (48). Later studies provided further evidence for discrimination between food and symbiont bacteria in the sponge *A. aerophoba* (49). The absence of dTDP-L-rhamnose and GDP-D-rhamnose O antigens on the LPS of “*Ca. Synechococcus spongiarum*” (implied by the absence of the biosynthetic genes) may provide this symbiont with resistance to host phagocytosis by a lack of recognition of these symbionts as consumable bacteria. However, rather than the “masking” being achieved by a capsule covering the element that is recognized by the phagocytic sponge cells, as previously suggested (48), it would be achieved by the lack of the element itself. Interestingly, mutants of the freshwater *S. elongatus* PCC7942 deficient in the synthesis of the O antigen were found to resist amoebal grazing (50). Further experimental work is necessary to test this hypothesis. Mutations in the genes involved in the production and transport of dTDP-L-rhamnose were found to also be responsible for phage resistance in the marine organism *Synechococcus* sp. strain WH7803 (51), suggesting that the lack of this O antigen in “*Ca. Synechococcus spongiarum*” may also provide protection from cyanophages, which could be particularly enriched by the sponge pumping activity. The lack of O antigen in free-living cyanobacteria does not decrease cell yield but rather promotes autoflocculation (51). This is a characteristic that may not concern the life of a symbiont that is embedded in the sponge mesohyl matrix but is selected against in free-living *Synechococcus*, as it may cause sinking to nonphotic zones.

**Divergent genomic features of the “*Ca. Synechococcus spongiarum*” species group.** The methionine salvage pathway (MSP), which recycles methionine from 5-methylthioadenosine (52), was only partially present in all four “*Ca. Synechococcus spongiarum*” genomes. Possibly, methionine is obtained from external sources, such as the host or sponge heterotrophic microbial consortium. The greater lack of MSP genes in SH4 may be due to the lower draft genome completeness compared with the other three genomes. Alternatively, SH4 may undergo genome reduction at a higher rate. Moreover, the predicted sizes of the four genomes

varied between different clades between 16 and 25% with SH4 being the smallest. Different clades of “*Ca. Synechococcus spongiarum*” may follow different symbiotic trajectories and thus have different degrees of genomic streamlining and dependencies on their hosts.

Siderophores are low-molecular-weight compounds that have high affinity to Fe(III) and are secreted to the environment, where they bind iron and then get transported back into the cell. The transport of siderophores into the cell is an energy-dependent mechanism and can include TonB receptors. The transport component of ABC-type siderophore systems consists of an extracellular substrate binding protein, an integral membrane protein, and ATPase (ATP hydrolases) (53). Only one marine cyanobacterium, *Cyanobium* sp. strain PCC7002, has so far been reported to harbor the genes for siderophore synthesis and transport (54). This iron uptake system is more distributed among cyanobacteria that are phylogenetically distant from “*Ca. Synechococcus spongiarum*,” such as the freshwater cyanobacteria *Synechococcus* sp. strain JA23 and *Synechocystis* sp. strain PCC6803, while the free-living *Synechococcus/Prochlorococcus* subclade phylogenetically closest to “*Ca. Synechococcus spongiarum*” lacks the ability for siderophore transport (54). All four “*Ca. Synechococcus spongiarum*” genomes harbored a gene coding for a membrane iron receptor likely related to siderophores (COG1629). However, only SP3 and 15L had all components of an active ABC-type iron transport system related to siderophores. Either SH4 and 142 could use a nonactive transport for the siderophores, or their COG1629 may sense a different type of available iron.

Eukaryotic-type domains are a common feature in microbial sponge symbionts (23, 24) and while ankyrin domain proteins are a typical genomic signature of sponge symbionts, the number of proteins with other eukaryotic-type domains (e.g., TPR and LRR) varied between the “*Ca. Synechococcus spongiarum*” phylogenotypes, with SP3 showing notably higher numbers of proteins with TPR domains than 142 and 15L, whereas 142 had more LRR-containing proteins (COG4886). It is tempting to speculate that the varying number of proteins containing LRR and TPR domains is a type of host-specific fingerprint and that each symbiont will harbor a certain combination of proteins with eukaryotic-type domains according to its specific host. Further research on the role of these domains is required, and more genomes of sponge-associated bacteria from the same sponge species need to be sequenced in order to provide support for this hypothesis.

As mentioned above, the presence of CRISPRs in cyanobacteria phylogenetically affiliated with the *Synechococcus/Prochlorococcus* clade was surprising. While the genome of strain 142 has two large CRISPR-Cas modules, the genomes of SH4, 15L, and SP3 harbored CRISPR-associated proteins dissociated from CRISPR regions (SH4 and 15L) or only CRISPR-associated proteins (SP3). Alternative antiviral defense mechanisms may also exist in “*Ca. Synechococcus spongiarum*.” For example, SH4 has two unique endonucleases (COG2810 and COG3587). Bacteriophage resistance mechanisms such as restriction-modification systems or genes preventing phage attachment to the cell surface could also be alternative immune system features (55). The latter mechanism may relate to the earlier-mentioned lack of a typical *Synechococcus* O antigen on the “*Ca. Synechococcus spongiarum*” LPS. The high level of intraspecies differences among antiviral mechanisms could also be related to the differences among the four host sponges. Parameters like water pumping behavior and

**TABLE 2** Functions enriched and depleted in “*Ca. Synechococcus spongiarum*” compared to members of the closely related free-living marine *Synechococcus/Prochlorococcus* subclade

Function	Context or interpretation (reference[s])
<b>Enriched</b>	
Recombination and repair	Insertion of mobile DNA into chromosomes (23, 24)
Transposable insertion elements	Horizontal gene transfer (32)
Eukaryotic-type domains	Ankyrin repeat domains possibly obligatory feature for sponge symbionts (37)
CRISPR-Cas system	Selective pressure to acquire phage resistance (higher exposure to viruses) (23, 24, 28)
ABC-type iron transport system	Retained ancestral function (lost in free-living subclade) (53, 54)
<b>Depleted</b>	
Cell wall biogenesis	Symbiotic minimalism (2)
Signal transduction mechanism	Symbiotic minimalism (39, 40)
Transcriptional regulation and (post)translational modification	Symbiotic minimalism (27)
ABC-type phosphate transport	Symbiotic minimalism (27)
Carbohydrate transport and metabolism and subunits of cytochrome <i>c</i>	Symbiotic minimalism (27)
Biosynthesis of LPS O antigen	Defense against phagocytosis by the sponge and anti-phage defense (45, 51)
Antioxidant enzymes	Reduced light radiation in sponge tissue (44)
Peptides of photosystem II and carotenoid biosynthesis	More stable light environment in the sponge tissue (2, 41, 43)
Methionine salvage pathway	Methionine obtained from external sources (52)

different exposure of the symbionts to the permanent incoming water of the hosts can reduce or increase the exposure to foreign DNA and phages. Different members of “*Ca. Synechococcus spongiarum*” may also be exposed to diverse persistent virus types, owing to biogeographic location. A pattern of maintaining “old” CRISPR sequences against persistent or reemerging viruses has been described (38). Thus, “*Ca. Synechococcus spongiarum*” intraspecies genomic divergence could potentially be affected by localized virus-host coevolution.

**Summary.** We have shown that besides almost identical 16S rRNA gene sequences, the “*Ca. Synechococcus spongiarum*” group possesses a number of intraspecies genomic differences, including those in genome size, gene content, immune system features, methionine *de novo* synthesis patterns, and eukaryotic-type proteins (LRR and TPR). However, ankyrin repeats seem to be conserved and common among different microbial phyla found in different sponge species and geographic locations, suggesting that ankyrin domain proteins are likely involved in the general recognition of sponge bacterial symbionts.

Table 2 summarizes the enriched and depleted functions in the genomes of “*Ca. Synechococcus spongiarum*” compared to the phylogenetically closest free-living cyanobacteria. The significantly higher proportion of the “replication, recombination and repair” (L) COG class matches well with findings from earlier metagenomic studies and likely also relates to horizontal gene transfer. The lower proportion of the COG class “signal transduction mechanisms” (T) may reflect a more stable environment inside the host than in the surrounding seawater.

The lack of biosynthesis genes for dTDP-L-rhamnose in “*Ca. Synechococcus spongiarum*” will affect the type of O antigen of its LPS. This may represent a novel mechanism for the discrimination of cyanobacterial symbionts from food cyanobacteria in sponges and for the resistance to phage attack in a niche likely characterized by high phage pressure. Further investigation to provide experimental evidence for these hypotheses is under way.

## MATERIALS AND METHODS

**Sampling and cell separation.** Sponge samples were collected by scuba diving: *Theonella swinhoei* from the Gulf of Aqaba, Red Sea, Israel, on

31 July 2012, *Ircinia variabilis* from Achziv nature marine reserve, Mediterranean Sea, Israel, on 5 May 2013, and *Aplysina aerophoba* from Piran, Mediterranean Sea, Slovenia, on 7 May 2013. Sponge samples were collected in compliance with permits 2012/38390 and 2013/38920 from the Israel Nature and National Parks Protection Authority. *T. swinhoei* and *I. variabilis* were transported on ice to the laboratory (IUI-Eilat and University of Haifa, respectively) for further processing (*T. swinhoei* within 20 min, *I. variabilis* within 2 h). *A. aerophoba* was transported to the laboratory (University of Würzburg, Germany) in natural seawater at ambient temperature; sponges were kept in a seawater aquarium and sampled within 1 week. Only cortex tissue was utilized for microbial cell enrichment from *I. variabilis* and *A. aerophoba*, while both cortex and endosome were used for *T. swinhoei*. Microbial cell enrichment was attained by a series of filtration and centrifugation steps (23) followed by DNA extraction for *T. swinhoei* and *I. variabilis*. For *A. aerophoba*, differential centrifugation (56) was utilized in combination with a cell sorting approach; fluorescence-activated cell sorting (FACS) was used targeting phycoerythrin and chlorophyll *a* autofluorescence of “*Ca. Synechococcus spongiarum*,” excited with a 488-nm laser. An enrichment of ca.  $10^6$  target cells was created. Two microliters of the enrichment was used as the template for multiple displacement amplification (MDA) using the REPLI-g single-cell kit (Qiagen) and following the manufacturer’s protocol using half the amount of any given reagent (25- $\mu$ l final reaction volume).

**Shotgun sequencing, assembly, and taxonomic binning.** The MDA product of the *A. aerophoba* symbiont was sequenced on an Illumina HiSeq2000 platform (150-bp paired-end reads). Sequences were quality filtered, and 3 Gbp were used for *de novo* assembly with SPAdes (57). The binning software CONCOCT (58) was used at default settings for decontamination. The phylogeny of the bins was determined with PhyloSift (59). rRNA genes were identified with rRNAPrediction (60) in the whole data set and added to the bin.

Metagenomic shotgun sequencing libraries of *T. swinhoei* and *I. variabilis* microbiomes were sequenced on the Illumina HiSeq2000 platform (100-bp paired-end reads). This generated ~13 and 7 Gb of sequence (from *I. variabilis* and *T. swinhoei*, respectively) with an average insert size of ~170 bp. FASTQ files were generated using CASAVA 1.8. Sequence quality was assessed, and low-quality reads ( $q = 3$ ) were trimmed using the FASTX toolkit 0.0.13.2 ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)). The sequence data set was assembled *de novo* using IDBA\_ud version 1.1.0 (61) with a kmer range of 50 to 70 and a step size of 5, following empirical tests. To genomically bin contigs, genes on contigs  $\geq 2$  kb long were predicted using Prodigal with the metagenome option (62). For each contig,



we determined the GC content, coverage, and the phylogenetic affiliation based on the best hit for each predicted protein in the Uniref90 database (September 2013) (63) following UBLAST searches (usearch64) (64). Contigs were assigned to genomes using these data, as well as emergent self-organizing map (ESOM)-based analysis of fragment tetranucleotide frequencies (65), as detailed by Handley et al. (66). To improve recovery of the highly abundant “*Ca. Synechococcus spongiarum*” in the *I. variabilis* sample, the genome was reassembled from the metagenome using Velvet with a kmer size of 59, expected genome kmer coverage of 707, and minimum and maximum cutoffs of 550 and 860, respectively. The assembled contigs were then rebinned as described above to verify correct genomic assignment.

**Whole-genome alignments and reordering of contigs.** Mauve version 20120303 (build 645) was used to align and rearrange the contigs of four “*Ca. Synechococcus spongiarum*” draft genomes, three derived from this study (SP3 from *T. swinhoei*, 142 from *I. variabilis*, and 15L from *A. aerophoba*) and one previously published (14). *Cyanobium gracile* PCC6377 was chosen as an alignment template for the initial reordering based on its high mean identity percentage in the amino acid matrix (see Table S1 in the supplemental material) and its close phylogenetic relatedness (Fig. 2). Percentage of aligned bases, locally collinear block numbers, and other parameters were obtained using Mauve Assembly Metrics (67). SP3 was chosen as the best candidate for the initial reordering because of its lowest number of contigs and highest completeness, and it was reordered using the Mauve reordering tool (67) against *C. gracile* PCC6307. In the final alignment, after 52 iterations, 1,082,448 bp of SP3 were aligned against 1,023,432 bp of *C. gracile* PCC6307. Thereafter, the contigs of SH4, 15L, and 142 were each rearranged against the reordered SP3 genome. To obtain regions of similarity and to inspect the structure of the reordered genomes, the reordered draft genome SP3 was BLAST searched against the reordered draft genomes 142, 15L, and SH4 using the BLASTn (Mega BLAST) and Artemis (<http://www.sanger.ac.uk/Software/ACT/>) genome alignment tools.

**ORF prediction, completeness estimation, and functional annotation.** Open reading frames (ORFs) in the four reordered “*Ca. Synechococcus spongiarum*” draft genomes and six closely related free-living cyanobacteria were identified and annotated with the classic RAST algorithm (22, 68). Clusters of orthologous groups (COGs) (69) and KEGGs (70) were annotated using RPSBLAST using the WebMGA annotation tool (71) with an e-value cutoff of 0.001.

Genome completeness was determined on the basis of 100 essential single-copy genes from a previously reported list of 111 genes (see essential.hmm in mmgenome pipeline on <https://github.com/MadsAlbertsen/>) (72), and 11 genes were omitted from the analysis based on their absence or multiple copy presence in the complete genomes of the free-living cyanobacteria analyzed here. Hmmssearch was conducted using hmmer v3.1b1 (<http://hmmer.janelia.org/>) to compare the amino acid sequences of all ORFs of each genome to the database of the 100 essential genes.

Genes found exclusively in all four genomes from “*Ca. Synechococcus spongiarum*” and absent from all six genomes from free-living cyanobacteria were obtained using the “calculate gene set feature” of the EDGAR platform, which is a reciprocal best-BLAST-hit approach (21). For this analysis, the four genomes from “*Ca. Synechococcus spongiarum*” were chosen as “included group” and the six genomes from free-living cyanobacteria as “excluded group.” Genes resulting from this analysis were BLAST searched against the COG database using WebMGA (71). The resulting COG annotations were compared to those obtained from ORFs derived from the six free-living cyanobacteria. For a gene to be considered unique to “*Ca. Synechococcus spongiarum*,” its COG annotation had to be absent from all six analyzed free-living cyanobacteria. Genes for which no COG annotation was available were annotated using the KEGG database as described above. STRING (25) (Search Tool for the Retrieval of Interacting Genes/Proteins) was used to predict interactions (coexpression, coexistence, cooccurrence, gene fusion, and neighborhood) between different COGs. eggNOG v. 3.0 (73) was used to obtain member lists of COGs

of interest. Protein domains were obtained using DELTA-BLAST against the refseq\_protein database on the NCBI website (<http://blast.be-md.ncbi.nlm.nih.gov/Blast.cgi>).

All *psb* genes were annotated using SEED with the exception of *psbN* and *psbY* in the genome of SP3, which were found using RPSBLAST against the KEGG database. *psbM* was obtained using BLASTn. Of 204 SEED annotated genes, 194 were identically annotated using RPSBLAST to search against the KEGG database.

**Statistical analyses of COG classes.** The relative abundance of each COG class and the comparison between COG classes in symbiotic versus free-living cyanobacteria was calculated using STAMP v2.0.9 (74). A heat map of the COG class abundance was created, and the average neighbor clustering (UPGMA) method was used to construct a dendrogram. Welch’s *t* test was used to determine significant differences in average relative abundances of COG classes between sponge-associated and free-living cyanobacteria. *P* values were corrected for multiple testing using the Bonferroni correction, and categories with corrected *P* values of <0.05 were considered significantly different.

**Comparative genomic analyses and phylogenomic tree.** Comparative genome analysis was performed using EDGAR, which uses an automatically calculated BLAST score ratio value cutoff for identifying conserved and unique gene sets among genome pairs based on bidirectional BLAST hit analyses (21). The pangenome of “*Ca. Synechococcus spongiarum*” was defined as the sum of all genes found in the four symbiotic genomes, while the core genome of “*Ca. Synechococcus spongiarum*” included only genes present in all four genomes. The amino acid identity matrix of the four symbiotic genomes was based on the mean percent identity values of genes from the core genome. The phylogenomic tree was constructed using EDGAR based on the core genome of the four “*Ca. Synechococcus spongiarum*” and 15 free-living cyanobacterial genomes, including two strains of *Synechococcus elongatus* as the outgroup. The phylogenomic analysis was performed in four steps: (i) amino acid sequences of core genes were identified by best reciprocal BLAST hits using *Cyanobium gracile* PCC6307 as the reference genome; (ii) they were aligned with their homologues using MUSCLE (75); (iii) the alignments of all genes were concatenated into a single alignment; and (iv) this alignment was used to reconstruct the phylogenomic tree using PHYLIP (21, 76). Bootstrap values were calculated from 100 replications using a Kimura distance matrix and neighbor-joining algorithm.

**CRISPR detection and analysis.** CRISPR arrays were predicted from the reordered contigs of the genomes with the online prediction tool CRISPRFinder (77) using the default setting, which identified confirmed and candidate CRISPRs. Only confirmed CRISPRs were used for further analysis. CRISPR-associated proteins were obtained using SEED and COG annotations. Whole-genome alignments using MAUVE (67) and Artemis (<http://www.sanger.ac.uk/Software/ACT/>) were used to obtain CRISPR-Cas modules. CRISPR-associated proteins were classified as described by Makarova and colleagues (29).

**16S rRNA genes and ITS and phylogenetic analyses.** For SP3 and 142, near-full-length 16S rRNA gene sequences were reconstructed from trimmed whole-genome shotgun reads using the EMIRGE (expectation maximization iterative reconstruction of genes from the environment) method (78). Genes were reconstructed over 100 iterations, following read mapping to a dereplicated version of the SILVA 108 small-subunit (SSU) rRNA database. EMIRGE-reconstructed sequences were verified by amplification of the 16S rRNA fragment and the 16S-23S internal transcribed spacer (ITS) region from DNA extracted from the specimens used for metagenome sequencing. The universal cyanobacterial forward primer 359F (79) and reverse primer 23S1R (80) were used for amplification, amplicons were cloned, and five clones each for SP3 and 142 were sequenced from both sides at Macrogen Europe (Amsterdam, Netherlands). For 15L, the 16S-23S ITS region was sequenced from MDA products (10 µl of a 1:2,500 dilution of MDA product was used as the template) using primers 1247f and 241r (81). The sequence was aligned and extended with the matching region in the whole-genome assembly of 15L.

The 16S-23S ITS region of SH4 was found in the assembled draft genome as a part of contig 126 (JENA01000091.1). The sequence alignment used to construct the phylogenetic tree included additional sequences that were downloaded from the NCBI nucleotide collection nonredundant database (<http://www.ncbi.nlm.nih.gov/>). Sequences were aligned using MUSCLE (75) with MEGA 6.0 (82). A maximum-likelihood tree based on distance estimates calculated by the Kimura 2-parameter substitution model with gamma-distributed rate variation (+G) and a proportion of invariant sites (+I) was constructed with MEGA 6.0. Phylogenetic robustness was inferred from 1,000 bootstrap replications.

**Nucleotide sequence accession numbers.** The 16S rRNA gene sequences of “*Ca. Synechococcus spongiarum*” 15L, SP3, and 142 were deposited in the NCBI GenBank database under accession numbers KP763586 (15L), KP792315 to KP792319 (142), and KP792320 to KP792324 (SP3). The draft genomes of the “*Ca. Synechococcus spongiarum*” SP3, 142, and 15L have been deposited at GenBank under accession numbers JXQG00000000, JXUO00000000, and JYFQ00000000, respectively.

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00391-15/-/DCSupplemental>.

Figure S1, PDF file, 0.3 MB.  
Figure S2, PDF file, 0.3 MB.  
Table S1, PDF file, 0.1 MB.  
Table S2, DOCX file, 0.02 MB.  
Table S3, DOCX file, 0.01 MB.  
Table S4, DOCX file, 0.01 MB.  
Table S5, DOCX file, 0.02 MB.  
Table S6, DOCX file, 0.01 MB.  
Table S7, DOCX file, 0.01 MB.  
Table S8, DOCX file, 0.01 MB.

## ACKNOWLEDGMENTS

Support for this study was provided by a USA-Israel Binational Science Foundation Young Investigator grant (BSF no. 4161011) to L.S. and a DOE Joint Genome Institute grant (CSP 1291) to U.H.; B.M.S. was supported by a grant of the German Excellence Initiative to the Graduate School of Life Sciences, University of Würzburg.

L.S. thanks Steve Giovannoni for productive discussions and useful suggestions during the project and the staff of the Inter-University Institute in Eilat, Israel, for their assistance. Sequencing was conducted at the Institute for Genomics and Systems Biology's Next Generation Sequencing Core (IGSB-NGS, ANL) and at the Joint Genome Institute (JGI) in Walnut Creek, California, USA. We thank Tanja Woyke and Frédéric Partensky for helpful discussions and advice. We also acknowledge the University of Chicago Research Computing Center for their support. Elena Burgsdorf is thanked for assistance with graphical design of figures.

## REFERENCES

- Rubin M, Berman-Frank I, Shaked Y. 2011. Dust- and mineral-iron utilization by the marine dinitrogen-fixing *Trichodesmium*. *Nat Geosci* 4:529–534. <http://dx.doi.org/10.1038/ngeo1181>.
- Larsson J, Nylander JA, Bergman B. 2011. Genome fluctuations in cyanobacteria reflect evolutionary, developmental and adaptive traits. *BMC Evol Biol* 11:187. <http://dx.doi.org/10.1186/1471-2148-11-187>.
- Crispim CA, Gaylarde CC. 2005. Cyanobacteria and biodeterioration of cultural heritage: a review. *Microb Ecol* 49:1–9.
- Steindler L, Huchon D, Avni A, Ilan M. 2005. 16S rRNA phylogeny of sponge-associated cyanobacteria. *Appl Environ Microbiol* 71:4127–4131. <http://dx.doi.org/10.1128/AEM.71.7.4127-4131.2005>.
- Thacker RW, Freeman CJ. 2012. Sponge-microbe symbioses: recent advances and new directions. *Adv Mar Biol* 62:57–111. <http://dx.doi.org/10.1016/B978-0-12-394283-8.00002-3>.
- Honda D, Yokota A, Sugiyama J. 1999. Detection of seven major evolutionary lineages in cyanobacteria based on the 16S rRNA gene sequence analysis with new sequences of five marine *Synechococcus* strains. *J Mol Evol* 48:723–739. <http://dx.doi.org/10.1007/PL00006517>.
- Usher KM. 2008. The ecology and phylogeny of cyanobacterial symbionts in sponges. *Mar Ecol* 29:178–192. <http://dx.doi.org/10.1111/j.1439-0485.2008.00245.x>.
- Burgsdorf I, Erwin PM, López-Legentil S, Cerrano C, Haber M, Frenk S, Steindler L. 2014. Biogeography rather than association with cyanobacteria structures symbiotic microbial communities in the marine sponge *Petrosia ficiformis*. *Front Microbiol* 5:529. <http://dx.doi.org/10.3389/fmicb.2014.00529>.
- Erwin PM, López-Legentil S, Turon X. 2012. Ultrastructure, molecular phylogenetics, and chlorophyll a content of novel cyanobacterial symbionts in temperate sponges. *Microb Ecol* 64:771–783. <http://dx.doi.org/10.1007/s00248-012-0047-5>.
- Usher KM, Kuo J, Fromont J, Sutton DC. 2001. Vertical transmission of cyanobacterial symbionts in the marine sponge *Chondrilla australiensis* (Demospongiae). *Hydrobiologia* 461:15–23. <http://dx.doi.org/10.1023/A:1012748727678>.
- Oren M, Steindler L, Ilan M. 2005. Transmission, plasticity and the molecular identification of cyanobacterial symbionts in the Red Sea sponge *Diacarnus erythraeus*. *Mar Biol* 148:35–41. <http://dx.doi.org/10.1007/s00227-005-0064-8>.
- Schmitt S, Angermeier H, Schiller R, Lindquist N, Hentschel U. 2008. Molecular microbial diversity survey of sponge reproductive stages and mechanistic insights into vertical transmission of microbial symbionts. *Appl Environ Microbiol* 74:7694–7708. <http://dx.doi.org/10.1128/AEM.00878-08>.
- Webster NS, Taylor MW, Behnam F, Lückner S, Rattei T, Whalan S, Horn M, Wagner M. 2010. Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environ Microbiol* 12:2070–2082. <http://dx.doi.org/10.1111/j.1462-2920.2009.02065.x>.
- Gao Z-M, Wang Y, Tian R-M, Wong YH, Batang ZB, Al-Suwailam AM, Bajic VB, Qian P-Y. 2014. Symbiotic adaptation drives genome streamlining of the cyanobacterial sponge symbiont “*Candidatus Synechococcus spongiarum*”. *mBio* 5:e00079–14. <http://dx.doi.org/10.1128/mBio.00079-14>.
- Hallam SJ, Konstantinidis KT, Putnam N, Schleper C, Watanabe Y, Sugahara J, Preston C, de la Torre J, Richardson PM, DeLong EF. 2006. Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc Natl Acad Sci U S A* 103:18296–18301. <http://dx.doi.org/10.1073/pnas.0608549103>.
- Kamke J, Szczyrba A, Ivanova N, Schwientek P, Rinke C, Mavromatis K, Woyke T, Hentschel U. 2013. Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. *ISME J* 7:2287–2300. <http://dx.doi.org/10.1038/ismej.2013.111>.
- Tian R-M, Wang Y, Bougouffa S, Gao Z-M, Cai L, Bajic V, Qian P-Y. 2014. Genomic analysis reveals versatile heterotrophic capacity of a potentially symbiotic sulfur-oxidizing bacterium in sponge. *Environ Microbiol* 16:3548–3561. <http://dx.doi.org/10.1111/1462-2920.12586>.
- Freeman CJ, Thacker RW. 2011. Complex interactions between marine sponges and their symbiotic microbial communities. *Limnol Oceanogr* 56:1577–1586. <http://dx.doi.org/10.4319/lo.2011.56.5.1577>.
- Freeman CJ, Thacker RW, Baker DM, Fogel ML. 2013. Quality or quantity: is nutrient transfer driven more by symbiont identity and productivity than by symbiont abundance? *ISME J* 7:1116–1125. <http://dx.doi.org/10.1038/ismej.2013.7>.
- Wu M, Eisen JA. 2008. A simple, fast, and accurate method of phylogenomic inference. *Genome Biol* 9:R151. <http://dx.doi.org/10.1186/gb-2008-9-10-r151>.
- Blom J, Albaum SP, Doppmeier D, Pühler A, Vorhölter F-J, Zakrzewski M, Goesmann A. 2009. Edgar: a software framework for the comparative analysis of prokaryotic genomes. *BMC Bioinformatics* 10:154. <http://dx.doi.org/10.1186/1471-2105-10-154>.
- Aziz RK, Bartels D, Best A, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75.
- Thomas T, Rusch D, DeMaere MZ, Yung PY, Lewis M, Halpern A, Heidelberg KB, Egan S, Steinberg PD, Kjelleberg S. 2010. Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. *ISME J* 4:1557–1567. <http://dx.doi.org/10.1038/ismej.2010.74>.



24. Fan L, Reynolds D, Liu M, Stark M, Kjelleberg S, Webster NS, Thomas T. 2012. Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proc Natl Acad Sci U S A* 109:E1878–E1887. <http://dx.doi.org/10.1073/pnas.1203287109>.
25. Snel B, Lehmann G, Bork P, Huynen MA. 2000. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res* 28:3442–3444. <http://dx.doi.org/10.1093/nar/28.18.3442>.
26. Snyder DS, Brahamsha B, Azadi P, Palenik B. 2009. Structure of compositionally simple lipopolysaccharide from marine *Synechococcus*. *J Bacteriol* 191:5499–5509. <http://dx.doi.org/10.1128/JB.00121-09>.
27. Nowack EC, Melkonian M, Glöckner G. 2008. Chromatophore genome sequence of *Paulinella* sheds light on acquisition of photosynthesis by eukaryotes. *Curr Biol* 18:410–418. <http://dx.doi.org/10.1016/j.cub.2008.02.051>.
28. Cai F, Axen SD, Kerfeld CA. 2013. Evidence for the widespread distribution of CRISPR-Cas system in the phylum Cyanobacteria. *RNA Biol* 10:687–693. <http://dx.doi.org/10.4161/rna.24571>.
29. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV. 2011. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 9:467–477. <http://dx.doi.org/10.1038/nrmicro2577>.
30. Bombar D, Heller P, Sanchez-Baracaldo P, Carter BJ, Zehr JP. 2014. Comparative genomics reveals surprising divergence of two closely related strains of uncultivated UCYN-A cyanobacteria. *ISME J* 8:2530–2542. <http://dx.doi.org/10.1038/ismej.2014.167>.
31. Liu M, Fan L, Zhong L, Kjelleberg S, Thomas T. 2012. Metaproteogenomic analysis of a community of sponge symbionts. *ISME J* 6:1515–1525. <http://dx.doi.org/10.1038/ismej.2012.1>.
32. Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw EA, Martin W, Esser C, Ahmadijeh N, Wiegand C, Madupu R, Beanan MJ, Brinkac LM, Daugherty SC, Durkin S, Kolonay JF, Nelson WC, Mohamoud Y, Lee P, Berry K, Young MB, Utterback T, Weidman J, Niernman WC, Paulsen IT, Nelson KE, Tettelin H, O'Neill SL, Eisen JA. 2004. Phylogenomics of the reproductive parasite *Wolbachia pipiensis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol* 2:E69.
33. Moliner C, Fournier P-E, Raoult D. 2010. Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. *FEMS Microbiol Rev* 34:281–294. <http://dx.doi.org/10.1111/j.1574-6976.2010.00209.x>.
34. Smillie CS, Smith MB, Friedman J, Cordero OX, David LA, Alm EJ. 2011. Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* 480:241–244. <http://dx.doi.org/10.1038/nature10571>.
35. Evdokimov AG, Anderson DE, Routzahn KM, Waugh DS. 2001. Unusual molecular architecture of the *Yersinia pestis* cytotoxin YopM: a leucine-rich repeat protein with the shortest repeating unit. *J Mol Biol* 312:807–821. <http://dx.doi.org/10.1006/jmbi.2001.4973>.
36. Marino M, Braun L, Cossart P, Ghosh P. 1999. Structure of the InlB leucine-rich repeats, a domain that triggers host cell invasion by the bacterial pathogen *L. monocytogenes*. *Mol Cell* 4:1063–1072. [http://dx.doi.org/10.1016/S1097-2765\(00\)80234-8](http://dx.doi.org/10.1016/S1097-2765(00)80234-8).
37. Nguyen MT, Liu M, Thomas T. 2014. Ankyrin-repeat proteins from sponge symbionts modulate amoebal phagocytosis. *Mol Ecol* 23:1635–1645. <http://dx.doi.org/10.1111/mec.12384>.
38. Weinberger AD, Sun CL, Pluciński MM, Deneff VJ, Thomas BC, Horvath P, Barrangou R, Gilmore MS, Getz WM, Banfield JF. 2012. Persisting viral sequences shape microbial CRISPR-based immunity. *PLOS Comput Biol* 8:e1002475. <http://dx.doi.org/10.1371/journal.pcbi.1002475>.
39. Tripp HJ, Bench SR, Turk KA, Foster RA, Desany BA, Niazi F, Affourtit JP, Zehr JP. 2010. Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium. *Nature* 464:90–94. <http://dx.doi.org/10.1038/nature08786>.
40. Kwan JC, Donia MS, Han AW, Hirose E, Haygood MG, Schmidt EW. 2012. Genome streamlining and chemical defense in a coral reef symbiosis. *Proc Natl Acad Sci U S A* 109:20655–20660. <http://dx.doi.org/10.1073/pnas.1213820109>.
41. Sveshnikov D, Funk C, Schröder WP. 2007. The PsbP-like protein (sl11418) of *Synechocystis* sp. PCC 6803 stabilises the donor side of photosystem II. *Photosynth Res* 93:101–109. <http://dx.doi.org/10.1007/s1120-007-9171-3>.
42. Usher KM, Kuo J, Fromont J, Toze S, Sutton DC. 2006. Comparative morphology of five species of symbiotic and non-symbiotic coccoid cyanobacteria. *Eur J Phycol* 41:179–188. <http://dx.doi.org/10.1080/09670260600631352>.
43. Meentemeyer M, Keren N, Ohad I, Pakrasi HB. 1999. The PsbY protein is not essential for oxygenic photosynthesis in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Physiol* 121:1267–1272. <http://dx.doi.org/10.1104/pp.121.4.1267>.
44. Latifi A, Ruiz M, Zhang C-C. 2009. Oxidative stress in cyanobacteria. *FEMS Microbiol Rev* 33:258–278. <http://dx.doi.org/10.1111/j.1574-6976.2008.00134.x>.
45. Lerouge I, Vanderleyden J. 2002. O-antigen structural variation: mechanisms and possible roles in animal/plant-microbe interactions. *FEMS Microbiol Rev* 26:17–47.
46. Pile A, Patterson M, Witman J. 1996. *In situ* grazing on plankton *Mycale lingua*. *Mar Ecol Prog Ser* 141:95–102. <http://dx.doi.org/10.3354/meps141095>.
47. Wilkinson CR. 1978. Microbial associations in sponges. III. Ultrastructure of the in situ associations in coral reef sponges. *Mar Biol* 49:177–185. <http://dx.doi.org/10.1007/BF00387117>.
48. Wilkinson CR, Garrone R, Vacelet J. 1984. Marine sponges discriminate between food bacteria and bacterial symbionts: electron microscope radioautography and in situ evidence. *Proc R Soc London Ser B Biol Sci* 220:519–528. <http://dx.doi.org/10.1098/rspb.1984.0018>.
49. Wehr M, Steinert M, Hentschel U. 2007. Bacterial uptake by the marine sponge *Aplysina aerophoba*. *Microb. Ecol* 53:355–365. <http://dx.doi.org/10.1007/s00248-006-9090-4>.
50. Simkovsky R, Daniels EF, Tang K, Huynh SC, Golden SS, Brahamsha B. 2012. Impairment of O-antigen production confers resistance to grazing in a model amoeba-cyanobacterium predator-prey system. *Proc Natl Acad Sci U S A* 109:16678–16683. <http://dx.doi.org/10.1073/pnas.1214904109>.
51. Marston MF, Pierciey FJ, Shepard A, Gearin G, Qi J, Yandava C, Schuster SC, Henn MR, Martiny JB. 2012. Rapid diversification of co-evolving marine *Synechococcus* and a virus. *Proc Natl Acad Sci U S A* 109:4544–4549. <http://dx.doi.org/10.1073/pnas.1120310109>.
52. Albers E. 2009. Metabolic characteristics and importance of the universal methionine salvage pathway recycling methionine from 5'-methylthioadenosine. *IUBMB Life* 61:1132–1142. <http://dx.doi.org/10.1002/iub.278>.
53. Köster W. 2001. ABC transporter-mediated uptake of iron, siderophores, heme and vitamin B 12. *Res Microbiol* 152:291–301. [http://dx.doi.org/10.1016/S0923-2508\(01\)01200-1](http://dx.doi.org/10.1016/S0923-2508(01)01200-1).
54. Hopkinson BM, Morel FM. 2009. The role of siderophores in iron acquisition by photosynthetic marine microorganisms. *Biometals* 22:659–669. <http://dx.doi.org/10.1007/s10534-009-9235-2>.
55. Stoddard LI, Martiny JB, Marston MF. 2007. Selection and characterization of cyanophage resistance in marine *Synechococcus* strains. *Appl Environ Microbiol* 73:5516–5522. <http://dx.doi.org/10.1128/AEM.00356-07>.
56. Fieseler I, Quaiser A, Schleper C, Hentschel U. 2006. Analysis of the first genome fragment from the marine sponge-associated, novel candidate phylum Poribacteria by environmental genomics. *Environ Microbiol* 8:612–624. <http://dx.doi.org/10.1111/j.1462-2920.2005.00937.x>.
57. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477.
58. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. 2014. Binning metagenomic contigs by coverage and composition. *Nat Methods* 11:1144–1150. <http://dx.doi.org/10.1038/nmeth.3103>.
59. Darling AE, Jospin G, Lowe E, Matsen FA, Bik HM, Eisen JA. 2014. PhyloSift: phylogenetic analysis of genomes and metagenomes. *PEER J* 2:e243. <http://dx.doi.org/10.7717/peerj.243>.
60. Huang Y, Gilna P, Li W. 2009. Identification of ribosomal RNA genes in metagenomic fragments. *Bioinformatics* 25:1338–1340. <http://dx.doi.org/10.1093/bioinformatics/btp161>.
61. Peng Y, Leung HC, Yiu SM, Chin FY. 2012. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28:1420–1428. <http://dx.doi.org/10.1093/bioinformatics/bts174>.

62. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <http://dx.doi.org/10.1186/1471-2105-11-119>.
63. Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. 2007. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* 23:1282–1288. <http://dx.doi.org/10.1093/bioinformatics/btm098>.
64. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <http://dx.doi.org/10.1093/bioinformatics/btq461>.
65. Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP, Banfield JF. 2009. Community-wide analysis of microbial genome sequence signatures. *Genome Biol* 10:R85. <http://dx.doi.org/10.1186/gb-2009-10-8-r85>.
66. Handley KM, VerBerkmoes NC, Steefel CI, Williams KH, Sharon I, Miller CS, Frischkorn KR, Chourey K, Thomas BC, Shah MB, Long PE, Hettich RL, Banfield JF. 2013. Biostimulation induces syntrophic interactions that impact C, S and N cycling in a sediment microbial community. *ISME J* 7:800–816. <http://dx.doi.org/10.1038/ismej.2012.148>.
67. Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14:1394–1403. <http://dx.doi.org/10.1101/gr.2289704>.
68. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2013. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res* 42: D206–D214. <http://dx.doi.org/10.1093/nar/gkt1226>.
69. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. <http://dx.doi.org/10.1186/1471-2105-4-41>.
70. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Res* 32: D277–D280. <http://dx.doi.org/10.1093/nar/gkh063>.
71. Wu S, Zhu Z, Fu L, Niu B, Li W. 2011. WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics* 12:444. <http://dx.doi.org/10.1186/1471-2164-12-444>.
72. Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol* 31:533–538.
73. Powell S, Szklarczyk D, Trachana K, Roth A, Kuhn M, Muller J, Arnold R, Rattei T, Letunic I, Doerks T, Jensen LJ, von Mering C, Bork P. 2012. eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res* 40:D284–D289. <http://dx.doi.org/10.1093/nar/gkr1060>.
74. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. 2014. STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 30: 3123–3124. <http://dx.doi.org/10.1093/bioinformatics/btu494>.
75. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <http://dx.doi.org/10.1093/nar/gkh340>.
76. Felsenstein J. 1995. PHYLIP: phylogeny inference package, version 3.57c. Seattle Department of Genetic. University of Washington.
77. Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 35:W52–W57. <http://dx.doi.org/10.1093/nar/gkm360>.
78. Miller CS, Baker BJ, Thomas BC, Singer SW, Banfield JF. 2011. EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol* 12:R44. <http://dx.doi.org/10.1186/gb-2011-12-5-r44>.
79. Nübel U, Garcia-Pichel F, Muyzer G. 1997. PCR primers to amplify 16S rRNA genes from cyanobacteria. *Appl Environ Microbiol* 63:3327–3332.
80. Iteman I, Rippka R, Tandeau De Marsac N, Herdman M. 2000. Comparison of conserved structural and regulatory domains within divergent 16S rRNA–23S rRNA spacer sequences of cyanobacteria. *Microbiology* 146:1275–1286.
81. Rocap G, Distel DL, Waterbury JB, Chisholm SW. 2002. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S–23S ribosomal DNA internal transcribed spacer sequences. *Appl Environ Microbiol* 68: 1180–1191. <http://dx.doi.org/10.1128/AEM.68.3.1180-1191.2002>.
82. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30: 2725–2729. <http://dx.doi.org/10.1093/molbev/mst197>.
83. Dufresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik BP, Paulsen IT, de Marsac NT, Wincker P, Dossat C, Ferriera S, Johnson J, Post AF, Hess WR, Partensky F. 2008. Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol* 9:R90. <http://dx.doi.org/10.1186/gb-2008-9-5-r90>.
84. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST Ring image generator (Brig): simple prokaryote genome comparisons. *BMC Genomics* 12:402. <http://dx.doi.org/10.1186/1471-2164-12-402>.