



Life at Home and on the Roam: Genomic Adaptions Reflect the Dual Lifestyle of an Intracellular, Facultative Symbiont

Iliia Burgsdorf,^a  Kim M. Handley,^b Rinat Bar-Shalom,^a Patrick M. Erwin,^c Laura Steindler^a

^aDepartment of Marine Biology, Leon H. Charney School of Marine Sciences, University of Haifa, Haifa, Israel

^bSchool of Biological Sciences, The University of Auckland, Auckland, New Zealand

^cDepartment of Biology and Marine Biology, Centre for Marine Science, University of North Carolina—Wilmington, Wilmington, North Carolina, USA

ABSTRACT “*Candidatus Synechococcus feldmannii*” is a facultative intracellular symbiont of the Atlanto-Mediterranean sponge *Petrosia ficiformis*. Genomic information of sponge-associated cyanobacteria derives thus far from the obligate and extracellular symbiont “*Candidatus Synechococcus spongiarum*.” Here we utilized a differential methylation-based approach for bacterial DNA enrichment combined with metagenomics to obtain the first draft genomes of “*Ca. Synechococcus feldmannii*.” By comparative genomics, we revealed that some genomic features (e.g., iron transport mediated by siderophores, eukaryotic-like proteins, and defense mechanisms, like CRISPR-Cas [clustered regularly interspaced short palindromic repeats-associated proteins]) are unique to both symbiont types and absent or rare in the genomes of taxonomically related free-living cyanobacteria. These genomic features likely enable life under the conditions found inside the sponge host. Interestingly, there are many genomic features that are shared by “*Ca. Synechococcus feldmannii*” and free-living cyanobacteria, while they are absent in the obligate symbiont “*Ca. Synechococcus spongiarum*.” These include genes related to cell surface structures, genetic regulation, and responses to environmental stress, as well as the composition of photosynthetic genes and DNA metabolism. We speculate that the presence of these genes confers on “*Ca. Synechococcus feldmannii*” its facultative nature (i.e., the ability to respond to a less stable environment when free-living). Our comparative analysis revealed that distinct genomic features depend on the nature of the symbiotic interaction: facultative and intracellular versus obligate and extracellular.

IMPORTANCE Given the evolutionary position of sponges as one of the earliest phyla to depart from the metazoan stem lineage, studies on their distinct and exceptionally diverse microbial communities should yield a better understanding of the origin of animal-bacterium interactions. While genomes of several extracellular sponge symbionts have been published, the intracellular symbionts have, so far, been elusive. Here we compare the genomes of two unicellular cyanobacterial sponge symbionts that share an ancestor but followed different evolutionary paths—one became intracellular and the other extracellular. Counterintuitively, the intracellular cyanobacteria are facultative, while the extracellular ones are obligate. By sequencing the genomes of the intracellular cyanobacteria and comparing them to the genomes of the extracellular symbionts and related free-living cyanobacteria, we show how three different cyanobacterial lifestyles are reflected by adaptive genomic features.

KEYWORDS *Petrosia ficiformis*, *Synechococcus feldmannii*, *Synechococcus spongiarum*, comparative genomics, endosymbionts, intracellular bacteria, metagenome assembled genomes, sponge, symbiosis

Citation Burgsdorf I, Handley KM, Bar-Shalom R, Erwin PM, Steindler L. 2019. Life at home and on the roam: genomic adaptations reflect the dual lifestyle of an intracellular, facultative symbiont. *mSystems* 4:e00057-19. <https://doi.org/10.1128/mSystems.00057-19>.

Editor Seth Bordenstein, Vanderbilt University

Copyright © 2019 Burgsdorf et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

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

Received 29 January 2019

Accepted 2 April 2019

Published 7 May 2019

Sponges are considered one of the earliest-branching multicellular animals (Meta-zoa) (1, 2). They inhabit marine and freshwater environments (3–5) and form intimate symbiotic interactions with complex communities of more than 60 phyla of bacteria, including cyanobacteria (6). The latter are a very diverse group with enormous influence on many processes at global scales, due to their ability to perform oxygenic photosynthesis (7–9). The two major groups of cyanobacteria have pronounced differences in their taxonomic affiliations, genome sizes and contents, morphological phenotypes, and developmental capacities of the cell (10–14). The first group, clade 1 (10), includes the highly diverse filamentous cyanobacteria that have dominated marine and freshwater benthic environments for more than 2,300 million years (11). The second group, clade 2 (10), includes marine unicellular cyanobacteria that originated much later (1,000 to 542 million years ago [Mya]) (11) and contributed to the emergence of metazoans, including the early-branching sponges (15), following the “Neoproterozoic oxygenation event” (11, 16–20). Modern planktonic cyanobacteria related to clade 2 consist of the recently proposed genus *Parasynechococcus* (12–14) and its sister clade, *Prochlorococcus*.

Sponge-associated cyanobacteria are polyphyletic, implying that they derived from multiple independent symbiotic events (21, 22). Unicellular cyanobacteria are the most commonly reported and are widespread in sponges (21), in particular “*Candidatus* *Synechococcus spongiarum*,” which consists of at least 12 different clades and is found in many different sponge species (23). A less common cyanobacterial symbiont, first defined as *Aphanocapsa feldmannii* (24) and later named “*Candidatus* *Synechococcus feldmannii*” (25), is a symbiont in the sponge species *Petrosia ficiformis* (Poiret, 1789) (21), which also harbors a dense and diverse microbiome (26). One study also reported the presence of “*Ca. Synechococcus feldmannii*” in *Ircinia variabilis*; however, the identification was based solely on morphology (24). Sequencing of the 16S-23S rRNA internal transcribed spacer regions from six *I. variabilis* specimens from Spain resulted in the identification of “*Ca. Synechococcus spongiarum*” and not “*Ca. Synechococcus feldmannii*” (27).

While “*Ca. Synechococcus spongiarum*,” like most sponge symbionts, is extracellular and vertically transmitted (via “leaky” vertical transmission) (28–31), “*Ca. Synechococcus feldmannii*” is thought to be transferred horizontally, as sponge oocytes of *P. ficiformis* were shown to lack symbiotic bacteria (32). In general, environmental acquisition of symbionts in *P. ficiformis* is supported by known biogeographic influences on the composition of its sponge-associated microbial community (26). Unlike “*Ca. Synechococcus spongiarum*,” “*Ca. Synechococcus feldmannii*” is localized inside specialized host cells called bacteriocytes (32, 33). The scenario of a facultative and horizontally transmitted intracellular symbiont (“*Ca. Synechococcus feldmannii*”) and an obligate and vertically transmitted extracellular symbiont (“*Ca. Synechococcus spongiarum*”) is peculiar. In other symbioses (e.g., insects with bacteria), symbionts present in bacteriocytes are usually linked to obligate relationships, where the symbiont relies on host-based mechanisms for transmission (34). In contrast, the obligate and vertically transmitted cyanobacterial symbiont “*Ca. Synechococcus spongiarum*” is found extracellularly in the sponge host rather than in bacteriocytes. Extracellular, obligate, and vertically transmitted symbionts can be also found in insects (35), yet they are understudied compared to their intracellular counterparts. While genome-based studies of extracellular and (primarily) vertically transmitted sponge symbionts have occupied the focus of published studies, including the cyanobacterium “*Ca. Synechococcus spongiarum*” (36, 37), genomes of bacteriocyte-associated sponge symbionts have so far been neglected. The characterization of facultative, intracellular sponge symbiont genomes will contribute to our understanding of the impact of endo-cellularity and transmission mode on bacterial genome evolution.

Here, we report the first genomes for the facultative intracellular species “*Ca. Synechococcus feldmannii*.” We compare these genomes with those from the extracellular, obligate symbiont species “*Ca. Synechococcus spongiarum*” and with free-

TABLE 1 Genomic information for the three “*Ca. Synechococcus feldmannii*” assemblies 277cV, 277cl, and 288cV

Assembly	Result for assembly		
	277cV	277cl	288cV
Genome size (Mb)	2.5	1.9	2.2
Avg GC content (%)	62.9	64.7	64
Completeness (%)	94.1	86.1	87.8
Contamination (%)	0.0	0.3	0.0
Heterogeneity (%)	0.0	0.0	0.0
No. of:			
Scaffolds	100	214	125
ORFs (Prodigal)	2,506	2,137	2,158
COGs ^a	1,084	1,018	1,008
SEED functions ^a	856	810	811
ORFs (RAST)	2,349	1,942	1,907
Ratio to pangenome (%)	80.3	66.4	65.2
No. of SNPs ^b	4,467	1,980	11,826
Nonsynonymous SNPs (%) ^b	73.5	66.5	77.5
Genes with SNPs (%) ^b	14.1	17.6	82.5
No. of scaffolds with SNPs ^b	100 (all)	210 (out of 214)	125 (all)
No. of SNPs ^c	1,349 for 277cl, 8,789 for 288cV		6,301 for 277cl
Nonsynonymous SNPs (%) ^c	73.7 for 277cl, 78.6 for 288cV		77.8 for 277cl

^aTotal number of unique COG and SEED annotations.

^bIllumina reads were mapped to the assembly.

^cArtificial reads produced from another genome were mapped to the assembly. The source of the artificial reads is given.

living cyanobacterial counterparts to reveal evidence of genomic adaptations to the different symbiotic lifestyles.

RESULTS

Genome recovery. Three draft genomes of “*Ca. Synechococcus feldmannii*” (277cV, 277cl, and 288cV) were obtained from two specimens of the Mediterranean sponge *P. ficiformis* (277 and 288). The pangenome and core genome of the three “*Ca. Synechococcus feldmannii*” specimens consisted of 2,924 genes and 1,338 genes, respectively. The core genome of the three “*Ca. Synechococcus feldmannii*” and six “*Ca. Synechococcus spongiarum*” specimens consisted of only 353 genes. “*Ca. Synechococcus feldmannii*” possesses relatively high GC content and small genome size similar to “*Ca. Synechococcus spongiarum*” and members of *Parasynechococcus* (36). Additional information on the “*Ca. Synechococcus feldmannii*” genomes is provided in Table 1.

Intraspecific genomic diversity. The three genomes of “*Ca. Synechococcus feldmannii*” showed similar gene architectures for homologous regions. However, each assembly contained different gene composition and single-nucleotide polymorphism (SNP) variations. A total of 1,349 SNPs were detected between the 277cV and 277cl genomes derived from the same sponge specimen (Table 1). The number of SNPs between the 277cV and 288cV, which was derived from a different *P. ficiformis* specimen, was higher: 8,789. Mapping of raw Illumina reads to the 277cV and 277cl genomes showed lower intragenomic variability (14.1% and 17.6% of genes had SNPs) than the 288cV genome (82.5%). However, the percentages of nonsynonymous mutations related to SNPs within the coding regions were relatively constant, ranging between 66.5% and 77.5% (Table 1). The genes containing SNPs within genome 277cV included all genes harboring fibronectin type III (FN3) and ankyrin (ANK) domains, 13 out of 14 genes coding for proteins with the leucine-rich repeat (LRR) domains, and 9 out of 10 CRISPR-Cas (clustered regularly interspaced short palindromic repeats-associated proteins)-related genes in the genome 277cl.

“*Ca. Synechococcus feldmannii*” and “*Ca. Synechococcus spongiarum*” are *Parasynechococcus*-like sponge-associated cyanobacteria. According to a phylogenomic analysis of 24 genomes (including 3 genomes of “*Ca. Synechococcus feldma-*

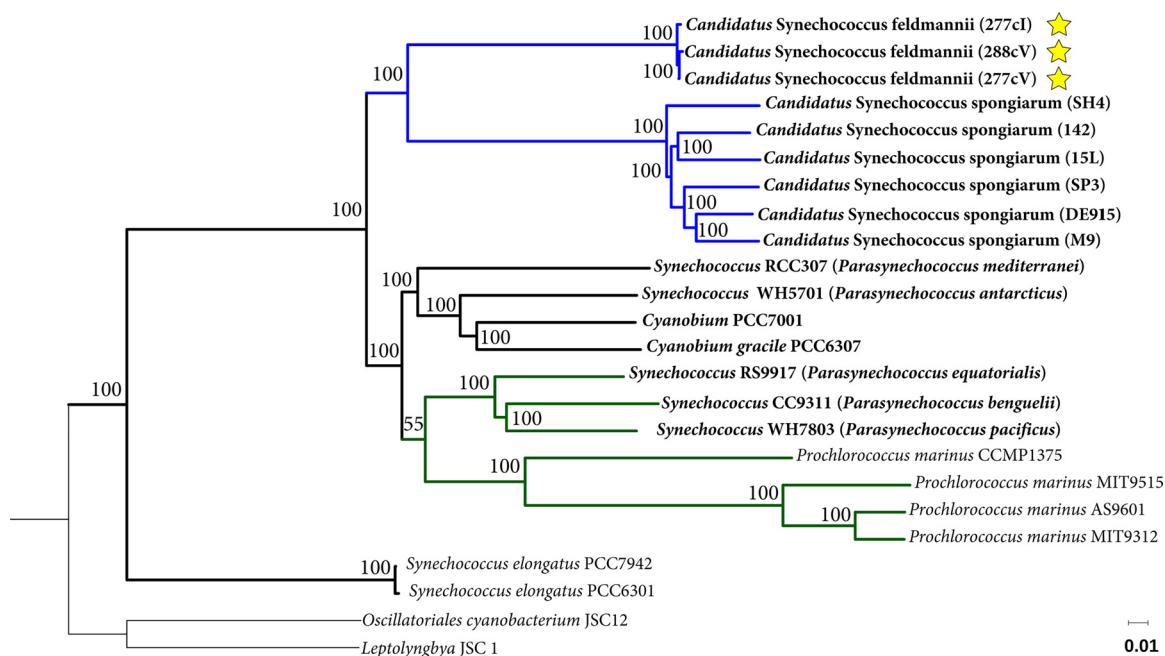


FIG 1 Concatenated phylogenetic core genome tree calculated by iterative pairwise comparison of genomes of the cyanobacteria analyzed here. Bootstrap values at branch nodes derive from 500 replications. The names of the here analyzed genomes are represented in bold. The genomes of “*Ca. Synechococcus feldmannii*” are marked with a star. Symbiotic cyanobacteria and the marine *Parasynechococcus/Prochlorococcus* subclade are marked with the blue and green branches, respectively. Cyanobacteria belonging to clade 2 (10) are marked with bold branches. *Oscillatoriales cyanobacterium* JSC12 and *Leptolyngbya* sp. strain JSC 1 belonging to clade 1 (10) were used as an outgroup for tree rooting.

nii,” 6 genomes of “*Ca. Synechococcus spongiarum*” and 15 genomes of free-living cyanobacteria) based on 135 core genes common to all genomes, “*Ca. Synechococcus feldmannii*” is most closely related to “*Ca. Synechococcus spongiarum*” (Fig. 1). These symbionts, together with *Prochlorococcus* and small-celled and mostly marine unicellular *Synechococcus* and *Synechococcus elongatus*, belong to cyanobacterial clade 2 and cluster separately from *Synechococcus* within cyanobacterial clade 1, characterized by a larger genome size and diverse phenotype (10). To address this phylogenetic characteristic of the group of closely related unicellular *Synechococcus* species, a proposal has recently been put forward to rename the group *Parasynechococcus* (13, 14). Thus, we refer to the symbionts as *Parasynechococcus*-like sponge-associated cyanobacteria.

The worldwide distribution of *Parasynechococcus*-like sponge-associated cyanobacteria. We analyzed the distribution and abundance of “*Ca. Synechococcus feldmannii*” and “*Ca. Synechococcus spongiarum*” using the Sponge Microbiome Project data set (6), which is part of the Earth Microbiome Project (EMP [www.earthmicrobiome.org]) and includes 2,882 sponge samples from 235 sponge species, 308 seawater samples, 54 marine sediment samples, and 1 algal tissue sample, collected from 37 countries. “*Ca. Synechococcus feldmannii*” and “*Ca. Synechococcus spongiarum*” were most closely affiliated (100% identity) to operational taxonomic units (OTUs) OTU0000398 and OTU0000007, respectively. “*Ca. Synechococcus feldmannii*” was found to be significantly enriched only in *P. ficiformis* (M. Britstein, C. Cerrano, L. Zoccarato, I. Burgsdorf, N. J. Kenny, A. Riesgo, M. Lalar, and L. Steindler, submitted for publication), establishing “*Ca. Synechococcus feldmannii*” as an intracellular symbiont that is highly specific to a single sponge species. In contrast, “*Ca. Synechococcus spongiarum*” was significantly enriched in 28 different sponge species (see Fig. S1 in the supplemental material) sampled in 21 (out of 37) countries around the globe. Nine sponge species significantly enriched in “*Ca. Synechococcus spongiarum*” were previously found to have photosynthetic activity based on measurements of photosynthetic quantum yield by pulse amplitude-modulated (PAM) fluorometry (38). Conversely, all

the PAM-negative sponge species lacked significant enrichment in “*Ca. Synechococcus spongiarum*.” Interestingly, “*Ca. Synechococcus spongiarum*” was relatively abundant in three samples of *P. ficiformis* from São Miguel Island, Portugal (1.84 to 5.63%), while the same specimens lacked their typical symbiont, “*Ca. Synechococcus feldmannii*.”

The functional genomic repertoire of *Parasynechococcus*-like sponge-associated cyanobacteria. Nine symbiotic cyanobacteria were compared to 19 *Parasynechococcus* cyanobacteria (Table 1; see Data Set S1, sheet 1, in the supplemental material). The choice of free-living *Parasynechococcus* species used for comparative genomics here is explained in (see Text S1 in the supplemental material). The genes of three “*Ca. Synechococcus feldmannii*” and six “*Ca. Synechococcus spongiarum*” cyanobacteria (Data Set S1, sheet 1) were assigned to 1,378 clusters of orthologous groups (COGs) and to 1,008 SEED functional roles. In comparison, the free-living genomes analyzed here received a higher number of COG and SEED annotations per genome (see Fig. S2 in the supplemental material). Agglomerative hierarchical clustering based on COG and SEED functional categories grouped the symbiont genomes together and apart from the closest free-living cyanobacteria. Symbiotic and free-living cyanobacteria were characterized by different enrichments of specific functional categories compared to free-living counterparts (see Fig. S3 in the supplemental material). A significantly lower proportion of functional categories in symbiotic genomes was confirmed by a lower number of genes with both COG and SEED annotations (Fig. S3 and Text S1). However, when considering both absolute counts and similar function based on COG annotation, only two SEED categories were truly enriched in symbiotic genomes: DNA metabolism and iron acquisition and metabolism (Fig. S3).

Nonmetric multidimensional scaling (NMDS) based on 1,430 SEED functional categories produced four separate clusters: “*Ca. Synechococcus feldmannii*” genomes, “*Ca. Synechococcus spongiarum*,” *Parasynechococcus*, and *Parasynechococcus*-related *Cyanobium* (39) clades (Fig. 2). A total of 764 out of the 1,430 functional categories significantly correlated to NMDS coordinates ($P \leq 0.05$). Symbiotic genomes were negatively loaded on axis 1 (Fig. 2A and B). Among the functions that influenced the separation between the two symbiont types, we detected a group of functional roles that appear related to the free-living stage of “*Ca. Synechococcus feldmannii*” (Fig. 2C). The dual lifestyle of “*Ca. Synechococcus feldmannii*” (free-living and resident within the host) is likely a source of a mixed gene composition in this symbiont, where some metabolic pathways showed similarity to the obligate symbiont “*Ca. Synechococcus spongiarum*” and others showed similarity to the closest free-living cyanobacteria, as detailed below.

Sponge-specific features in the *Parasynechococcus*-like sponge symbionts. Despite the overall functional similarity, “*Ca. Synechococcus spongiarum*” and “*Ca. Synechococcus feldmannii*” shared only 21 out of 1,378 COGs (Data Set S1, sheet 2). Shared depleted and enriched functions likely reflect their adaptation to a common niche: the host sponge. Among the shared depleted functions were genes related to maltose/maltodextrin and manganese transport, DNA repair system UmuCD, and arsenate reductase (Data Set S1, sheet 3). Among shared enriched functions, we found genes related to (i) siderophore-mediated iron transport and (ii) defense mechanisms against invading genomes.

(i) Siderophore-mediated iron transport in sponges. Symbiotic cyanobacteria were found to be enriched in genes related to the transport of siderophores: COG0609, COG1629, and COG4558 (Data Set S1, sheet 2). Eight out of nine symbiotic genomes contained genes related to siderophore transport, while the complete pathway of siderophore-mediated iron transport was found in one genome of “*Ca. Synechococcus feldmannii*” (277cV) and two “*Ca. Synechococcus spongiarum*” genomes (M9 and SP3). An operon including the iron regulation gene *irpA* (COG3487) was found in all genomes of “*Ca. Synechococcus feldmannii*.” Interestingly, COG3487 was mostly present in *Alphaproteobacteria* and *Gammaproteobacteria* (including the genera *Pseudomonas*,

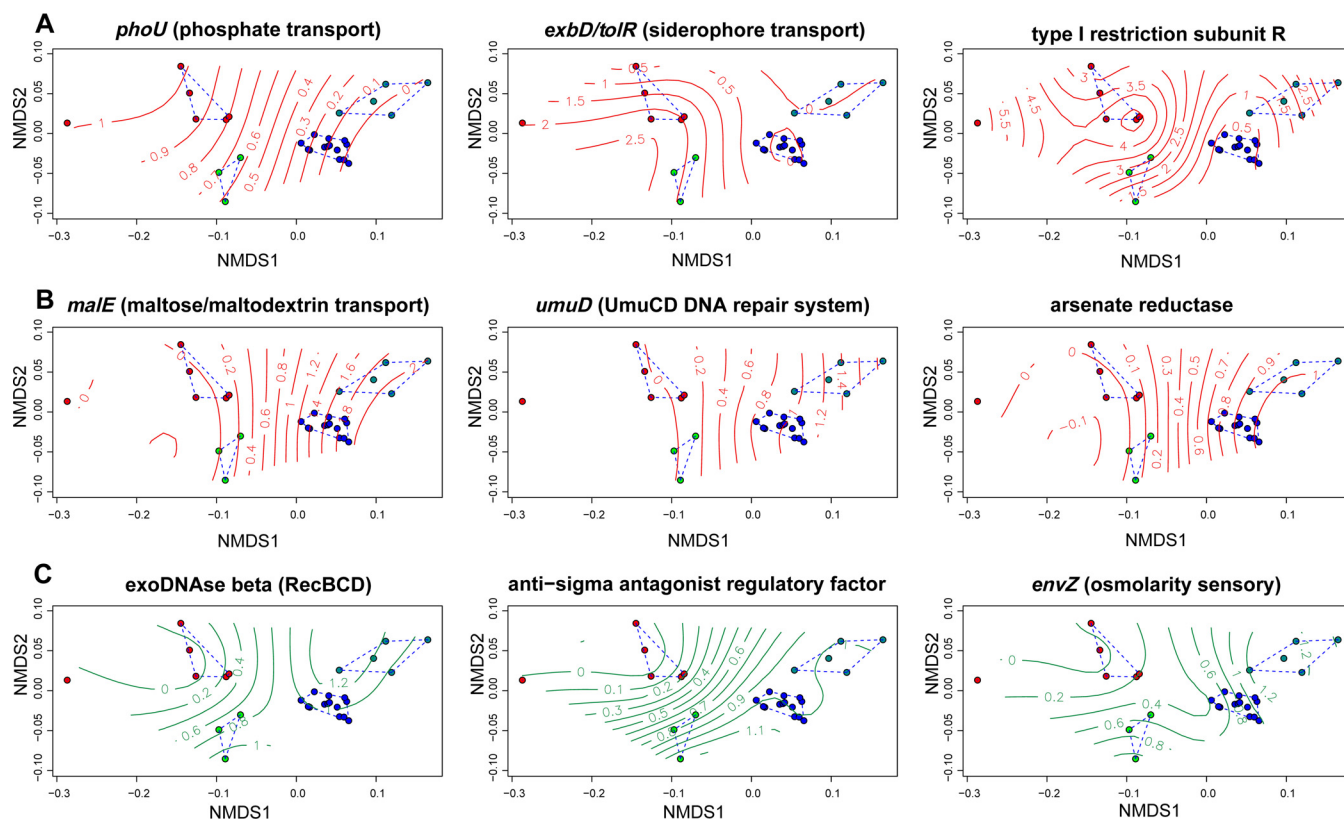


FIG 2 Nonmetric multidimensional scaling (NMDS) ordination plots on relative abundances of 1,430 SEED annotations of 28 cyanobacterial genomes, NMDS stress value = 0.058. Clusters of genomes produced are marked with a dashed blue line. Individual plots display overlaid smooth surfaces for significant ($P < 0.01$) representative functional roles (A) enriched in symbiotic cyanobacteria, (B) depleted in symbiotic cyanobacteria, and (C) common between “*Ca. Synechococcus feldmannii*” and free-living cyanobacteria. Numbers on the splines represent number of copies of the specific functional role. Red color splines represent distinctive features for symbiotic cyanobacteria. Green color splines represent common features for “*Ca. Synechococcus feldmannii*” and free-living cyanobacteria. “*Ca. Synechococcus spongiarum*,” “*Ca. Synechococcus feldmannii*,” *Parasynechococcus* clade, and *Parasynechococcus/Cyanobium* are marked with red, green, blue, and cyan dots, respectively.

Vibrio, and *Agrobacterium*) and, according to eggNOG, was absent from all members of the *Parasynechococcus/Prochlorococcus* clade.

(ii) Defense mechanisms against invading genomes. Out of the 21 common COGs, only COG1106, related to abortive bacterial infection (40), was present in all 9 genomes of sponge-associated cyanobacteria (Data Set S1, sheet 2), while among the free-living cyanobacteria, it was present only in cyanobacteria distantly related to the *Parasynechococcus*-like sponge symbionts (e.g., *Trichodesmium erythraeum* IMS101 and *Nostoc punctiforme* PCC73102). The total number and composition of restriction-modification system (RMS) genes in the symbiotic cyanobacteria, and also four members of the *Cyanobium* clade (e.g., *Parasynechococcus antarcticus*) and RS9917 (*Parasynechococcus equatorialis*), were higher and different, respectively, from those of the remaining 14 free-living cyanobacteria (Fig. 3). CRISPR-Cas is another type of defense mechanism against invading DNA. Eight and four CRISPR regions were found in the 277cV and 277cI “*Ca. Synechococcus feldmannii*” genomes, respectively. However, only 277cI contained a pronounced CRISPR-Cas-related region and harbored CRISPR-Cas-related genes (Fig. 4A). Genes related to CRISPR-Cas were also found in all six genomes of “*Ca. Synechococcus spongiarum*” and in 2 (WH8016 and WH8020) out of the 19 free-living *Parasynechococcus* clade genomes we analyzed (Fig. 4A). Surprisingly, 277cI shared seven CRISPR-Cas system eggNOG orthologs (NOGs) with the free-living WH8020 (Fig. 4A). In contrast, CRISPR-Cas-related genes found in “*Ca. Synechococcus spongiarum*” genomes resembled orthologs from other phyla or distantly related cyanobacteria (Fig. 4A). For example, *cse1* gene sequences found in three genomes of

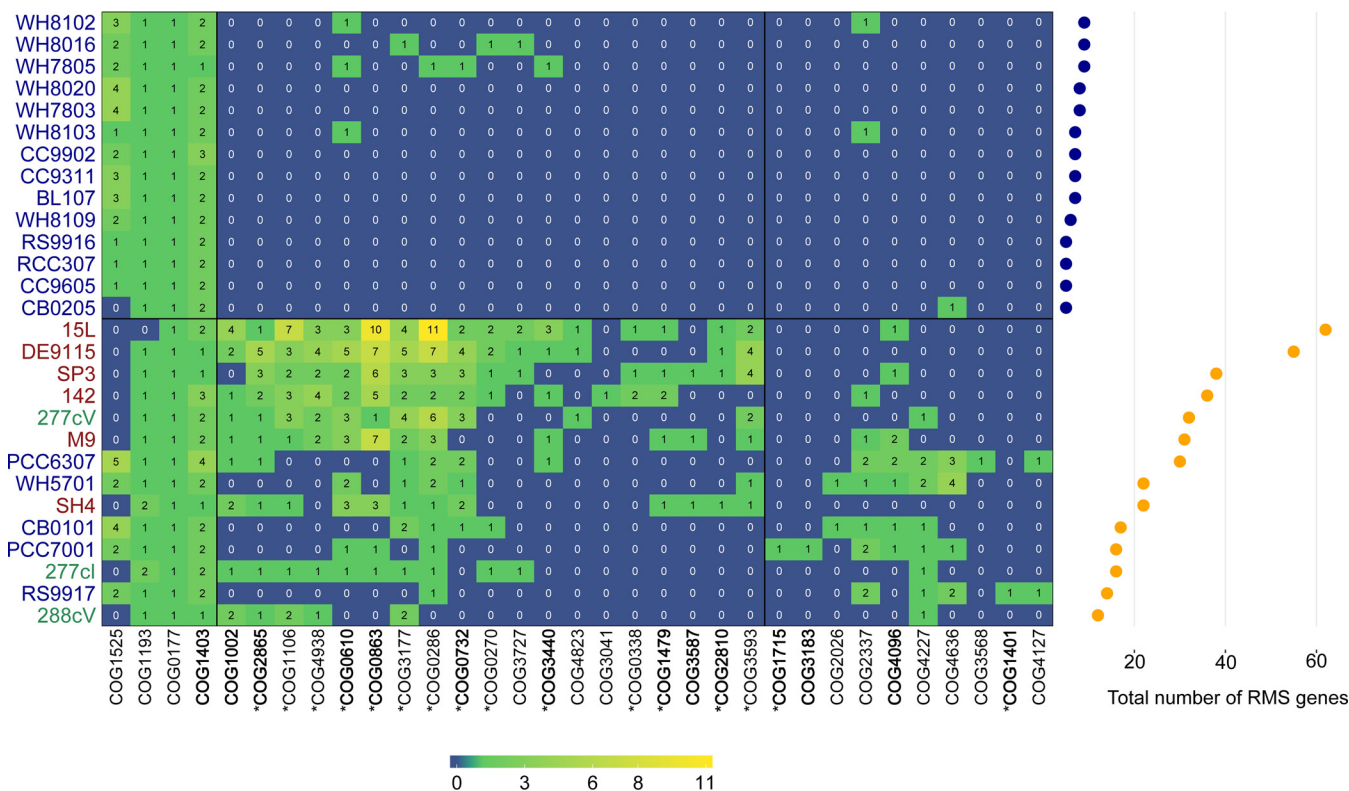


FIG 3 Heat map of absolute counts of genes from different COGs related to microbial defense systems related to “*Ca. Synechococcus feldmannii*” (green), “*Ca. Synechococcus spongiarum*” (red) and free-living bacteria (blue). The genomes were divided into two groups based on binary distance matrix. The total number of genes related to RMS is present on the right. All the genomes of “*Ca. Synechococcus spongiarum*” (red) are incapable to perform homologous recombination. COGs mentioned in Slaby et al. (77) as significantly enriched in genomes of sponge symbionts are marked with a star. COGs previously found as enriched in sponge-associated metagenomes in Horn et al. (76) are marked in bold.

“*Ca. Synechococcus spongiarum*” were most closely related to orthologs from *Gammaproteobacteria* (Fig. 4B).

Features potentially related to the free-living stage of “*Ca. Synechococcus feldmannii*.” “*Ca. Synechococcus feldmannii*” is hypothesized to be a facultative symbiont (32, 33) and thus to have a free-living stage that may share relevant adaptive genes with the members of the *Parasynechococcus* clade. Indeed, we find that “*Ca. Synechococcus feldmannii*” shares 99 COGs with 19 free-living counterparts (Data Set S1, sheet 4). The shared genes were related to (i) cell surface and motility, (ii) DNA metabolism, (iii) environmental stress and cell regulation, and (iv) photosynthesis (Data Set S1, sheet 3). The relatively higher number of shared COGs between “*Ca. Synechococcus feldmannii*” and free-living *Parasynechococcus* compared to those shared between “*Ca. Synechococcus feldmannii*” and “*Ca. Synechococcus spongiarum*” is discussed in Text S1.

(i) Cell surface and motility. Enzymes involved in the synthesis of the O antigen residue L-rhamnose, including dTDP-glucose pyrophosphorylase (COG1209 [EC 2.7.7.24]), dTDP-4-dehydrorhamnose 3,5-epimerase (COG1898 [EC 5.1.3.13]), and dTDP-4-dehydrorhamnose reductase (COG1091 [EC 1.1.1.133]), were annotated in 17 out of 19 *Parasynechococcus* genomes and in “*Ca. Synechococcus feldmannii*” 288cV. dTDP-4-dehydrorhamnose reductase (COG1091 [EC 1.1.1.133]) was also annotated in 277cV. This is in contrast to “*Ca. Synechococcus spongiarum*,” in which these genes were found to be missing (36, 37), possibly to avoid host predation. Unlike “*Ca. Synechococcus spongiarum*,” all three “*Ca. Synechococcus feldmannii*” and *Parasynechococcus* genomes possessed a motion-related pilus retraction ATPase *pilT* gene (COG2805).

(ii) DNA metabolism. All three genomes of “*Ca. Synechococcus feldmannii*” harbored small and large subunits of exonuclease VII (or exodeoxyribonuclease VII). The

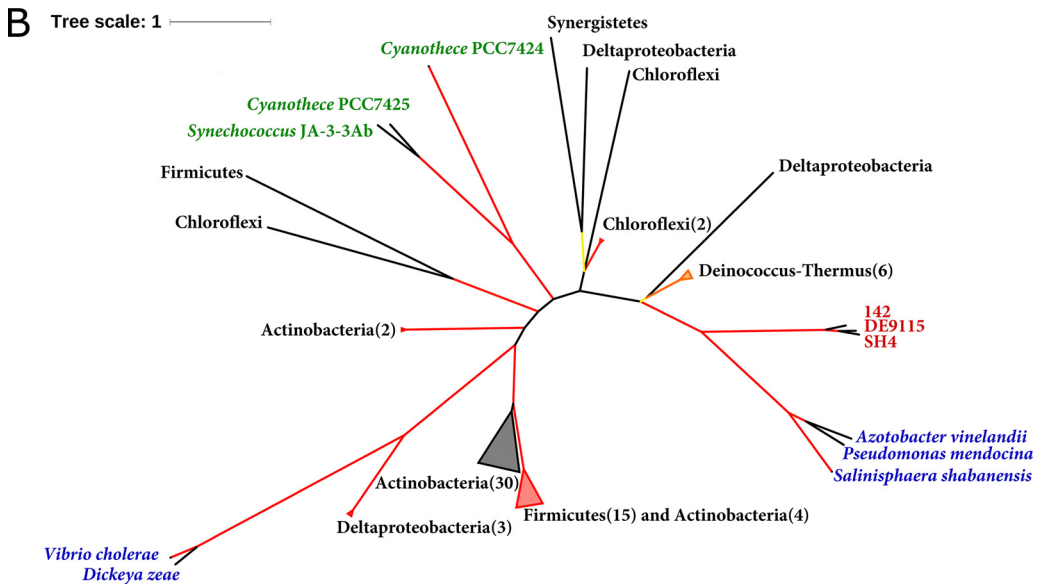
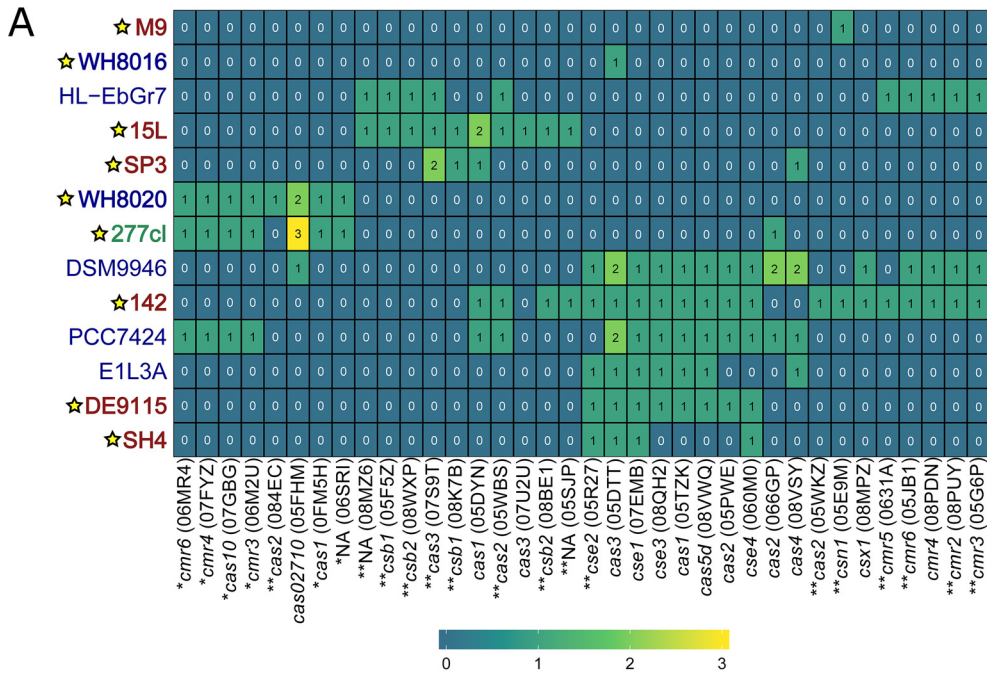


FIG 4 (A) Heat map of absolute counts of genes from different orthologous groups (NOGs) related to the CRISPR-Cas system. *Ca. Synechococcus feldmannii* (green), *Ca. Synechococcus spongiarum* (red), and free-living bacteria (blue) are compared. Only NOGs relevant for the genomes analyzed here (yellow star on the left) are shown. E1L3A represents *Salinisphaera shabanensis* E1L3A, HL-EbGr7 represents *Thioalkalivibrio sulfidiphilus* HL-EbGr7, PCC7424 represents *Cyanothecae* sp. strain PCC 7424, and DSM9946 represents *Meiothermus silvanus* DSM9946. NOGs exclusively found in cyanobacteria and not found in cyanobacteria are marked with one star and two stars, respectively. (B) Maximum likelihood phylogenetic tree of the protein Cse1 (07EMB). *Ca. Synechococcus spongiarum* is marked in red, *Gammaproteobacteria* in blue, and free-living cyanobacteria in green. Branches with similar taxonomy were collapsed, and numbers of collapsed orthologs are shown in parentheses. The branches were colored according to the bootstrap values, ranging from yellow (55%) to red (100%).

recBCD pathway (exonuclease V), including alpha, beta, and gamma subunits, was present in the 277cV and 288cV genomes, while 277cl lacked the gamma subunit (COG1330). Both exonucleases were absent in *Ca. Synechococcus spongiarum* genomes and were present in all *Parasynechococcus* genomes (Data Set S1, sheet 3).

(iii) Environmental stress and cell regulation. The spermidine synthesis pathway (COG0421 [EC 2.5.1.16] and COG1586 [EC 4.1.1.50]), important for survivability under

TABLE 2 Summary of ELP types in symbiotic cyanobacteria

ELP	Predicted interaction(s)	Interaction with host cells/tissue	Result for ELP types:	
			" <i>Ca. Synechococcus feldmannii</i> " (n = 3)	" <i>Ca. Synechococcus spongiarum</i> " (n = 6)
Ankyrin LRR	Avoidance of host predation Host specific	Archaeocyte Unknown	277cV All genomes	All genomes 4 genomes (except 15L and M9)
TPR Rhamnose-free O antigen Cadherin	Host specific Avoidance of host predation and phage resistance Adhesion to host cells	Unknown Archaeocyte Unknown	Not enriched 277cV and 288cV Enriched in 277cV and 288cV	Enriched in SP3 Absent Enriched in DE9115
FN3 ^a	Binding to host integrins and possible colonization	Extracellular matrix	277cV and 288cV	5 genomes (except M9)

^aThere were 3 to 11 FN3 proteins per genome (18 to 80 FN3 domains per genome) for "*Ca. Synechococcus feldmannii*" and 2 to 5 FN3 proteins per genome (2 to 12 FN3 domains per genome) for "*Ca. Synechococcus spongiarum*."

low temperatures and osmotic stress, was annotated in all three genomes of "*Ca. Synechococcus feldmannii*" and absent in "*Ca. Synechococcus spongiarum*" (Data Set S1, sheet 6). Anti-sigma and anti-sigma antagonist factors are adaptive gene expression regulators correlated to environmental changes (41). Anti-sigma antagonist (COG1366) and serine phosphatase RsbU, regulator of sigma subunit (COG2208), were among the 17 orthologous groups found present in all genomes of free-living cyanobacteria and lacking from "*Ca. Synechococcus spongiarum*" (141). Interestingly, eight of these orthologous groups, including COG1366 and COG2208, were shared between "*Ca. Synechococcus feldmannii*" and free-living members of the *Parasynechococcus* clade. Furthermore, "*Ca. Synechococcus feldmannii*," like members of the free-living *Parasynechococcus* analyzed here, harbored histidine kinase/phosphatase *envZ* and related transcription factor *ompR*, predictably lacking in "*Ca. Synechococcus spongiarum*" (Data Set S1, sheet 3).

(iv) Photosynthesis. Photosystem II *psbP* and *psbY* genes were absent in all six genomes of "*Ca. Synechococcus spongiarum*" and present in all three genomes of "*Ca. Synechococcus feldmannii*" (Data Set S1, sheet 5).

Unique features of "*Ca. Synechococcus feldmannii*." The peculiar lifestyle of "*Ca. Synechococcus feldmannii*," an intracellular symbiont that is horizontally transmitted, likely requires functions that differ from both its closest free-living relatives and from the extracellular, primarily vertically transmitted, symbiont "*Ca. Synechococcus spongiarum*." Fourteen COGs were exclusively found in two out of three genomes of "*Ca. Synechococcus feldmannii*" (Data Set S1, sheet 7). FimT (COG4970), related to pilin structure, was annotated in all three "*Ca. Synechococcus feldmannii*" genomes (Data Set S1, sheet 3). Additional DELTA-BLAST analysis confirmed the annotation of the FimT domain in COG4970. Only two genomes from the *Parasynechococcus* clade that we analyzed harbored COG4970. All "*Ca. Synechococcus spongiarum*" genomes lacked these genes.

ELPs. Eukaryotic-like proteins (ELPs) found in the different symbiont types and free-living cyanobacteria are summarized in Table 2. The "*Ca. Synechococcus feldmannii*" 277cV genome contained 11 proteins with 80 fibronectin type III (FN3) domains (Table 2; Data Set S1, sheet 8). Three proteins (277cV_123, 277cV_1068, and 277cV_2528) showed a mixed-domain architecture (Fig. 5A) and also contained cadherin (CAD), cadherin-like (CHDL), and autotransporter domains (Fig. 5A). A search for proteins with a domain architecture similar to that of 277cV_123 in the Conserved Domain Architecture Retrieval Tool (CDART) database revealed 15 proteins: 14 of them belong to the *Xanthomonadales* order of *Gammaproteobacteria*. Interestingly, one protein belonged to the cyanobacterial sponge symbiont "*Ca. Synechococcus spongiarum*" SP3: SP3_1976 (37) (Fig. 5B). For 277cV_1068, the proteins with the closest architecture belonged to the plant endosymbiotic nitrogen-fixing genus *Rhizobium*. The CHDL-containing protein 277cV_641 of "*Ca. Synechococcus feldmannii*" also harbored a

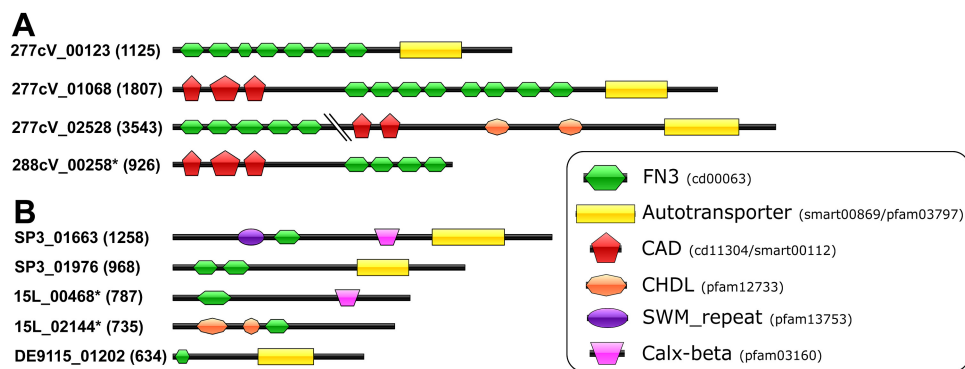


FIG 5 Domain architecture of FN3 proteins combined with additional types of domains derived from (A) “*Ca. Synechococcus feldmannii*” and (B) “*Ca. Synechococcus spongiarum*.” Names of proteins with an incomplete C terminal are marked with a star. The numbers of residues are written in parentheses.

CshA-type fibril repeat (CSHAF). Surprisingly, this protein architecture (CSHAF and CHDL) appears to be unique to “*Ca. Synechococcus feldmannii*.” CSHAF was also found in WH8102 (*Parasynechococcus marenigrum*) and WH7803 (*Parasynechococcus pacificus*); however, the domain composition of these proteins differed from 277cV.

Five out of six genomes of “*Ca. Synechococcus spongiarum*” also contained proteins with FN3 domains (Table 2). The “*Ca. Synechococcus spongiarum*” 15L, DE9115, and SP3 genomes had 5 proteins that contained other types of domains in addition to FN3 (Fig. 5 and Table 2). In contrast, the 19 free-living cyanobacteria analyzed here all lacked proteins containing FN3 domains (Table 2). A global search of proteins with FN3 domains within the entire phylum *Cyanobacteria* revealed 63 sequences that are not sponge associated; only 8 of them showed a mixed-domain architecture, like that found in the symbionts (Data Set S1, sheet 9). Among those 63 sequences, only one belonging to *Synechococcus* sp. strain GFB01 (out of more than 40 species of free-living cyanobacteria with FN3 domains) was found to be affiliated with clade 2 cyanobacteria (10), the same clade to which the *Parasynechococcus*-like sponge symbiotic cyanobacteria belong (Fig. 1). The two major taxonomic groups of free-living cyanobacteria that harbored FN3 domain proteins belonged to the *Nostocales* and *Oscillatoriales* orders (Data Set S1, sheet 9).

DISCUSSION

Different lifestyles result in different genomic adaptations. The cyanobacterial sponge symbiont “*Ca. Synechococcus feldmannii*” is exceptional in terms of its unique association with the sponge species *P. ficiformis*, combined with its horizontal acquisition and intracellular location (Fig. 6). In contrast, a stable mutualism characterized by vertical transmission is inherent for the widespread cyanobacterial symbiont “*Ca. Synechococcus spongiarum*” (43). The unique lifestyle of “*Ca. Synechococcus feldmannii*” may reflect a more ancestral symbiotic relationship between sponges and cyanobacteria, before a stable mutualistic state was established.

“*Ca. Synechococcus spongiarum*” is vertically transmitted, so it likely does not need to survive a free-living state (outside the host). Contrarily, “*Ca. Synechococcus feldmannii*” is acquired from the environment at each sponge generation and thus must be able to survive a free-living stage and recolonize a sponge host. Furthermore, the extracellular “*Ca. Synechococcus spongiarum*” is exposed to sponge archaeocytes (sponge phagocytizing cells), while the intracellular “*Ca. Synechococcus feldmannii*” is protected from sponge archaeocytes by a physical barrier (the bacteriocyte membrane). We found that these differences in lifestyle are likely reflected in several genetic features of these two symbiont types (Fig. 6). We previously suggested that the typical O antigen of *Parasynechococcus*, composed of L-rhamnose (44), may be recognized by sponge archaeocytes as a signal for phagocytosis, while the lack of an L-rhamnose-based O antigen in the symbiont “*Ca. Synechococcus spongiarum*” (according to the

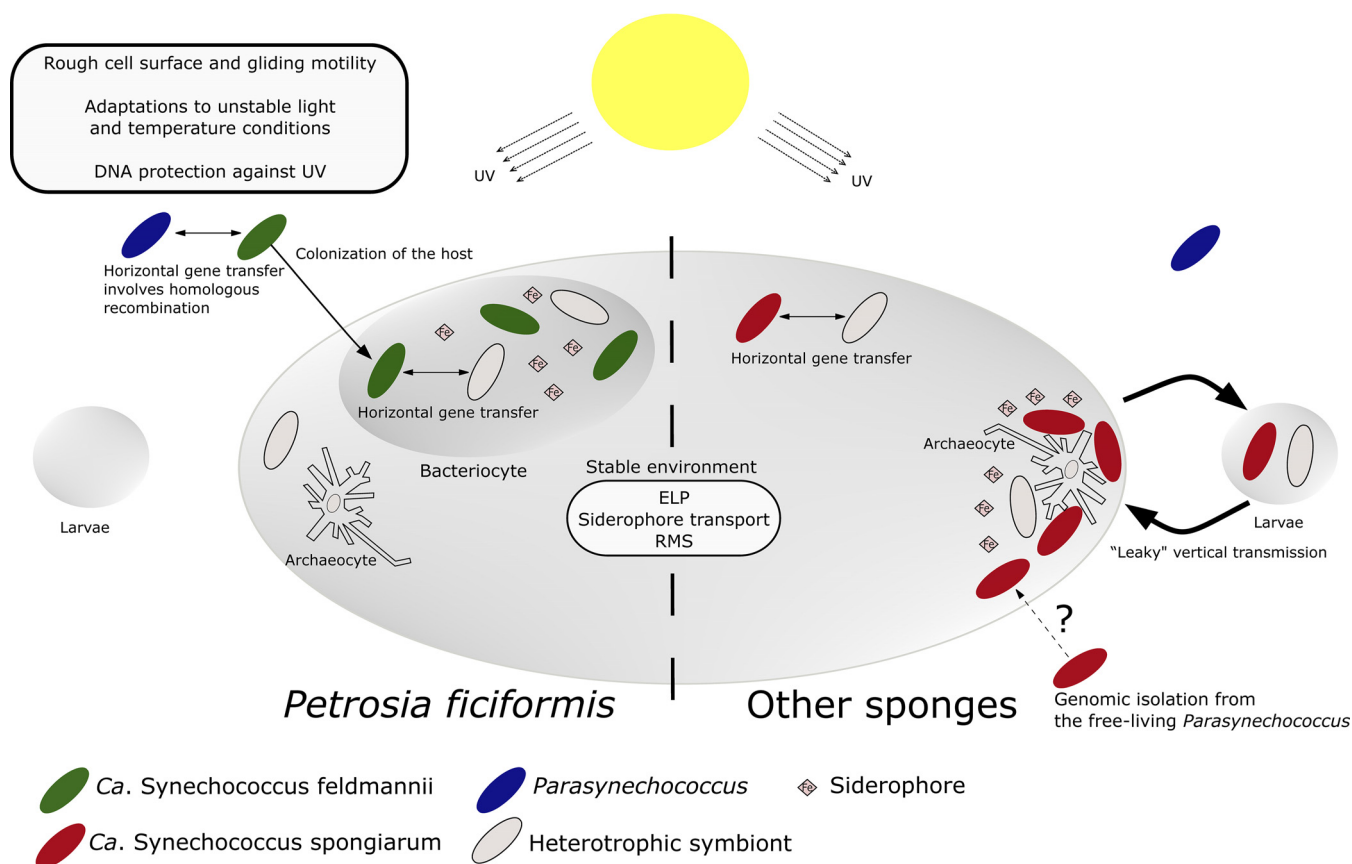


FIG 6 Schematic representation of genetic features that reflect different symbiotic lifestyles inherent to “*Ca. Synechococcus feldmannii*” and “*Ca. Synechococcus spongiarum*.” These differences incorporate types of symbiont transmission, degree of genomic isolation toward free-living cyanobacteria, iron metabolism, and cell surface properties related to interaction with the host archaeocytes. RMS, restriction-modification systems; ELP, eukaryotic-like proteins.

missing biosynthetic pathway for this sugar) may enable this symbiont to avoid host predation (37). Free-living cyanobacteria, on the other hand, must maintain the L-rhamnose-based O antigen, in order to float in the water column (because O antigen mutants of free-living cyanobacteria sink [45]). As expected, the facultative symbiont “*Ca. Synechococcus feldmannii*,” which is predicted to have a free-living stage and thus a requirement to float, would need to maintain biosynthesis genes to produce the L-rhamnose-based O antigen. “*Ca. Synechococcus feldmannii*,” when found inside the host sponge, is suggested to use a different mechanism to protect itself from sponge archaeocyte phagocytosis: the compartmentalization inside bacteriocyte cells. We further speculate that “*Ca. Synechococcus feldmannii*” fails to colonize other sponge species, as in their mesohyl it will be vulnerable to host predation.

Common adaptations to the sponge environment. Iron is one of the vital factors for cyanobacterial metabolism due to its high requirements by the photosynthetic apparatus (46). Most marine cyanobacteria, including the *Parasynechococcus/Prochlorococcus* clade, lack production and transport of siderophores and rather use cell surface reduction of Fe(III) to the soluble Fe(II) (47). However, “*Ca. Synechococcus feldmannii*,” similar to “*Ca. Synechococcus spongiarum*” genomes, harbored a significantly higher number of genes related to iron acquisition, including siderophore transport. *Parasynechococcus*-like sponge symbionts lacked genes related to synthesis of siderophores and thus likely actively compete for siderophores from external sources (43). Discovery of the siderophore producer-consumer relations within the sponge-associated community will be an important step in understanding the network of chemical dependencies among the sponge community members. Interestingly, and dissimilar to “*Ca. Synechococcus spongiarum*,” “*Ca. Synechococcus feldmannii*” harbors

an operon that includes an *irpA*-like iron-regulated gene (COG3487) that is related to iron starvation in *Synechococcus* sp. strain PCC7942 (48–50). Changes in environmental iron conditions may play a role for “*Ca. Synechococcus feldmannii*” to sense whether it is found inside or outside the host and accordingly trigger gene regulation related to either the symbiotic or the free-living state.

ELPs: similarities and differences between the two symbiont types. One type of common ELP is ANK repeat proteins. Pathogenic proteins containing ANK repeats were shown to prevent digestion of the pathogens in the vacuoles of the host immune system cells (51, 52). A similar mechanism is also used by nonpathogenic bacteria: for example, a very high number of ANK-containing proteins in insect-associated *Wolbachia* were linked to the interaction with the host (53, 54). FN3 domain proteins are another type of ELP linked to adhesion of the bacterium to host tissues. The adhesin protein FliA (containing FN3 domains) was found in the pathogen *Campylobacter jejuni* and appeared to be essential for its attachment to epithelial host cells (55). Moreover, functional genomic and exoproteomic analyses of the S-layer-forming probiotic *Lactobacillus* revealed the presence of extracellular adhesins containing FN3 domains (56–58). ELPs such as ANK, CAD, FN3, and LRR domains were previously reported to be enriched in various (meta)genomic analyses of sponge symbionts (36, 37, 59–64), and also in the *Parasynecococcus*-like symbionts of sponges analyzed here. However, the different symbiotic properties of “*Ca. Synechococcus feldmannii*” and “*Ca. Synechococcus spongiarum*” likely influence the arsenal of ELP genes each harbor. The extracellular nature of “*Ca. Synechococcus spongiarum*” probably necessitates the observed higher diversity of genes with ANK repeats due to its direct exposure to the phagocytosing archaeocyte host cells (65, 66), while the requirement of “*Ca. Synechococcus feldmannii*” for sponge cell colonization explains the observed higher number of FN3 and CAD domains, which have adhesion properties. The enrichment of FN3 domains in sponge-associated bacteria was proposed to be involved in adhesion to the host cells by attachment to glycoprotein and structural proteins (60, 67), as well as cadherin domains, in adhesion to host cells and host colonization (60). Interestingly, we found that the expression of a sponge protein containing an FN3-like domain was positively correlated to the acquisition of “*Ca. Synechococcus feldmannii*” by *P. ficiformis* (M. Britstein, personal communication). FN3 domains with mixed-domain architecture from “*Ca. Synechococcus feldmannii*” showed higher similarity to *Alpha*- and *Gamma*proteobacteria than to free-living cyanobacteria and thus may have been acquired by horizontal gene transfer (HGT) from co-occurring sponge symbionts.

A high percentage of nonsynonymous mutations among ELP genes in “*Ca. Synechococcus feldmannii*” may reflect directional selection on these genes for improved colonization of the host. For example, adaptive mutations in adhesin genes of *Escherichia coli* increased binding to the polysaccharide structures of the host (68). Alternatively, the high mutation rates on genes related to host interaction may relate to their location on the cell membrane. Genomic islands characterized by higher mutation and HGT rates were found to contain genes involved in the formation of extracellular structures, among diverse free-living cyanobacteria, and their acquisition is generally thought to serve as a defense system against phage pressure (69–72). Interestingly, genes related to the CRISPR-Cas system in “*Ca. Synechococcus feldmannii*” also had a high mutation rate characteristic of genomic islands.

Ecological isolation and genetic barriers in the two symbiont types. The majority of the free-living *Parasynecococcus* strains analyzed here do not possess pronounced genomic barriers and are characterized by a large number of HGT events (73–75). Conversely, the symbiont genomes were found to be enriched in restriction-modification systems (RMSs) and genes related to abortive infection, in accordance with what has previously been reported for (meta)genomes of sponge symbiotic bacteria (59, 60, 76–79). RMSs (80, 81) and abortive infection systems (40) are efficient and widespread defense mechanisms against invading DNA in various bacteria. These systems promote genomic diversity by isolation of different genotypes from the

genetic exchange with other lineages (80–83). The RMS (type I) allows bacteria to distinguish between self and foreign DNA according to methylation pattern. RMSs may thus promote HGT with distantly related bacteria of a similar methylation pattern, but may lead to genetic isolation toward closely related bacteria with different RMS repertoires (59, 60, 76). HGT can also occur by homologous recombination, involving *recBCD* (84, 85), which plays an important role in genome evolution of bacterial populations (86–89). RecBCD has multiple roles and can also function as a defense mechanism against invading DNA by degrading foreign DNA, if the crossover hot spot instigator (CHI) sequence is absent in the incoming DNA (80, 84). Earlier studies reported that the presence of a high number of RMS genes in bacterial genomes correlates with the absence of *recBCD* (80), which is also the case in “*Ca. Synechococcus spongiarum*.” Thus, “*Ca. Synechococcus spongiarum*” may exchange genetic material only with other nearby symbionts that have similar RMS patterns, while it may have lost the ability for homologous recombination with other cyanobacteria (5). “*Ca. Synechococcus feldmannii*,” on the other hand, maintains the *recBCD* genes and has a free-living stage. It may, therefore, encounter other cyanobacteria and be capable of exchange of genetic information with the free-living *Parasynechococcus* strains by homologous recombination. Alternatively, homologous recombination involving RecBCD and ExoVII may also be linked to protection against UV damage (85, 90, 91) that “*Ca. Synechococcus feldmannii*” might face during the free-living stage. In accordance with the above results, and based on gene homology, we hypothesize that “*Ca. Synechococcus feldmannii*” obtained its CRISPR-Cas system through HGT from a free-living *Parasynechococcus* strain, while “*Ca. Synechococcus spongiarum*” probably acquired it from a sponge-associated gammaproteobacterial symbiont.

“*Ca. Synechococcus feldmannii*”: special requirements to survive the free-living stage. The facultative nature of the association of “*Ca. Synechococcus feldmannii*” with its host sponge requires the ability to adapt to changing environments and may involve more capacity for gene regulation than that of the obligate symbiont “*Ca. Synechococcus spongiarum*.” Despite a lower number of histidine kinases (COG0642) and DNA-binding response regulators (COG0745), “*Ca. Synechococcus feldmannii*” indeed differed from “*Ca. Synechococcus spongiarum*” in its repertoire of genes involved in genetic regulation and transcription, sharing with free-living cyanobacteria several components that were found lacking in “*Ca. Synechococcus spongiarum*”: for example, genes *envZ* and *ompR*, known to control the osmoregulation of *Escherichia coli* (42, 92).

While both symbiont types had a streamlined genetic component of the photosynthesis apparatus, the composition of these genes in “*Ca. Synechococcus feldmannii*” was more similar to that of free-living cyanobacteria and included *psbP* and *psbY* genes, which were lacking in “*Ca. Synechococcus spongiarum*.” These genes were probably lost in “*Ca. Synechococcus spongiarum*,” as it is always found in a more light-stable sponge environment and is characterized by lack of competition with different cyanobacterial species (37, 93, 94). However, such genes may be required in “*Ca. Synechococcus feldmannii*” to survive light shifts during the time it is found outside the sponge. Similarly, the presence of *N*-acetylmuramic acid 6-phosphate etherase (*murQ*), which is used in the reutilization of degradation products of peptidoglycan under limited-light conditions (95), may be important for “*Ca. Synechococcus feldmannii*” when the symbiont is found in seawater at low light levels. “*Ca. Synechococcus spongiarum*” lost this function. Moreover, spermidine is one of the most abundant polyamines in cyanobacteria (96, 97) and was previously linked to the replacement of damaged proteins under “chill-light” conditions (i.e., low temperature in combination with light) and osmotic stresses in *Synechocystis* sp. strain PCC6803 (98, 99). In *Bacillus subtilis*, the sigma-regulated phosphatase RsbU is involved in response to environmental stresses, including blue light and osmolytes (100, 101). The complete loss of genes involved in spermidine biosynthesis and sigma regulation genes in “*Ca. Synechococcus spongiarum*” may be related to a stable environment inside the host, while the presence of this

biosynthetic pathway in “*Ca. Synechococcus feldmannii*” may be needed during the free-living stage.

Besides functional metabolic adaptations, “*Ca. Synechococcus feldmannii*” harbors cell surface elements that better resemble its free-living counterparts rather than the symbiont “*Ca. Synechococcus spongiarum*.” These include gliding motility-related *pilT* (102, 103) and pilus assembly-related *fimT* (104–106). The absence of these genes in “*Ca. Synechococcus spongiarum*” and their presence in “*Ca. Synechococcus feldmannii*” likely reflect the free-living stage of “*Ca. Synechococcus feldmannii*” and the need for host colonization, through increased contact with the host cells and motility toward the sponge.

In summary, we have shown that “*Ca. Synechococcus feldmannii*” combines features of planktonic and symbiotic picocyanobacteria. The facultative nature of its symbiosis with *P. fificiformis* would enable the host to select for the optimal symbionts according to the environmental conditions sensed. This would create competition among potential substrains of “*Ca. Synechococcus feldmannii*,” which are capable of exchanging genetic information with both open-ocean and symbiotic bacteria, promoting adaptive evolution.

MATERIALS AND METHODS

Sponge sampling, DNA isolation, and microbial DNA purification. Two (no. 277 and 288) *Petrosia fificiformis* specimens were collected by SCUBA in January 2014 at depths of 27.3 and 14.9 m, respectively, from the Achziv nature marine reserve, Mediterranean Sea, Israel. Sponges were collected in compliance with permit 40246/2014 from the Israel Nature and National Parks Protection Authority. Only cortex tissue was used for further DNA extraction. DNA was extracted as described earlier (107). The microbial DNA fraction was enriched using New England Biolab’s NEBNext microbiome DNA enrichment kit.

Shotgun sequencing, assembly, and binning. Genomic DNA was fragmented by sonication using Covaris S2 (Covaris, Woburn, MA). DNA libraries were prepared using the KAPA Hyper DNA library preparation kit with further Pippin Prep (Sage Scientific) size selection to 800 to 1,000 bp. Metagenomic shotgun libraries were sequenced on an Illumina NextSeq 500 platform (150-bp paired-end reads) in the DNA Services Facility at the University of Illinois at Chicago. Totals of 384.2 and 308.6 Gb of sequence for 277 and 288, respectively, were generated with mean insert sizes of 673 and 679.9 bp. Low-quality (minimum-quality threshold = 20) and short (minimum-length threshold = 50) reads and reads with ambiguous bases (“N”) were trimmed with the software sickle version 1.33, using a sliding-window approach (108). Sequence quality was evaluated using FastQC version 0.11.5 (109).

The 277cI genome was assembled *de novo* using IDBA-UD version 1.1.0 (110) (k-mer range = 40 to 80, step = 10). The 277cV and 288cV genomes were assembled from a subsampled version of the data set (10% of reads) using Velvet, with a k-mer size of 59, expected genome k-mer coverage of 41, and minimum cutoff of 15. Only scaffolds of ≥ 2 kb were used for genome binning. Genomes were first binned either based on visualization of a self-organizing map (ESOM) using tetranucleotide signatures as previously described (111) or on DNA fragment clustering via Barnes-Hut stochastic neighbor embedding (BH-SNE) in VizBin using pentanucleotide signatures (112). Second, assembly errors were rectified using REAPR version 1.0.18 (113). Third, completeness and contamination of the final bins were estimated with checkM version 1.0.7 (114). Then, protein sequences were obtained using Prodigal in metagenome mode (115). Finally, taxonomy affiliation of the predicted genes was obtained as detailed earlier (37).

ORF prediction, domain search, and functional annotation. Protein sequences were queried against the NCBI-CDD database using the Domain Enhanced Lookup Time Accelerated Basic Local Alignment Search Tool (Delta-BLAST, BLAST version 2.2.30+) and Reverse Position-Specific BLAST (RPSBLAST, BLAST version 2.2.30+), with E value cutoffs of 0.05 and 0.001, respectively. CAD and CHDL (cd11303, cd11304, PFAM00028, PFAM12733, and smart00736), FN3 (cd00063), and CSHAF (TIGR04225) domains were obtained from the Delta-BLAST results. The architecture of the relevant proteins was visualized with the IBS illustrator (116). COG annotation version 2014 (117) was assigned with Perl script `cdd2cog.pl` (<https://github.com/aleimba/bac-genomics-scripts/tree/master/cdd2cog>) using the RPS-BLAST results. The amino acid sequences identified were also searched against the NCBI-NR database with DIAMOND (blastp, sensitive) and assigned to SEED/Subsystems annotation (118) using MEGAN 6.11.4. Additional functional annotation of protein sequences was performed by eggNOG-mapper and eggNOG database version 4.5 (119, 120), using the bactNOG data set and HMMER data mapping mode. The eggNOG website (<http://eggnogdb.embl.de>) was also used to obtain member lists of relevant orthologous groups. The Conserved Domain Architecture Retrieval Tool (CDART) (121) (<https://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi>) was used to obtain proteins with similar domain architecture. The Integrated Microbial Genomes (IMG) (122) (<https://img.jgi.doe.gov/cgi-bin/mer/main.cgi>) and Pfam (123) (<http://pfam.xfam.org/>) databases were used to obtain proteins with FN3 domains among cyanobacteria.

Statistical analyses of the genome functional constituency. Bray-Curtis dissimilarities were calculated based on the SEED/Subsystems annotations using the `vegdist` function (vegan package) in R version 3.4.1. Multivariate nonmetric multidimensional scaling (NMDS) was created using the `metaMDS` function (vegan package in R) (124, 125). Clusters of the genomes were created using `hclust` (ward.D2

agglomeration method) and a cut height of 0.32 by the cutree function. Smooth surfaces were fitted using the thin plate spline method in the `ordisurf` function (`vegan` package). Heat maps were created using the `superheat` package (126) (Manhattan or binary distances, `ward.D2` agglomeration method for hierarchical clustering). The Wilcoxon test (`wilcox.test` in R, package `stats`) was used to determine significant differences in relative abundances of COG and SEED functional classes between sponge-associated and free-living cyanobacteria. *P* values were corrected for multiple testing using the Bonferroni correction (`p.adjust` function in R), and categories with corrected *P* values of <0.05 were considered significantly different.

Comparative genomic and phylogenomic analyses. The phylogenomic tree was constructed using EDGAR (127) based on the core genome as detailed by Burgsdorf et al. (37). The pangenomes of symbionts were obtained as the set of all genes in the relevant group of genomes (37). For these purposes, open reading frames (ORFs) in the three “*Ca. Synechococcus feldmannii*,” six “*Ca. Synechococcus spongiarum*,” and 15 free-living cyanobacteria were identified with the classic RAST algorithm (128, 129).

CRISPR detection and analysis. CRISPR arrays were predicted using CRISPRFinder (130). Only CRISPRs described as confirmed were used for the further analysis. CRISPR-associated proteins were annotated using `eggNOG-mapper`. Protein sequences of Cse1 (07EMB) were obtained from the `eggNOG` website and aligned using GUIDANCE2 (131–133) (<http://guidance.tau.ac.il/ver2/>) using the MAFFT algorithm and 100 iterations. Ambiguous regions were trimmed from the alignment by `trimAl` (134) using the `gappyout` method. A maximum likelihood tree of 320 positions was calculated by the `Le_Gascuel_2008` model (135), with gamma-distributed rate variation (1.4908), and a proportion of invariant sites (0.4688%) was constructed with MEGA 7.0.2 (136). Phylogenetic robustness was inferred from 100 bootstrap replications. The resulting tree was redrawn and annotated using `iTOL` (137).

Distribution and abundance of cyanobacteria in sponge versus environmental samples. To determine the distribution and abundance of “*Ca. Synechococcus spongiarum*,” we used the SMP data set (6, 138). 16S rRNA sequences were subjected to `blastn` 2.2.30+ (139) (E value threshold = 0.005) against the OTU representatives of the amplicon sequencing data set contains sponge and environmental samples (6). The OTUs with sequence identity of $\geq 99\%$ were used to determine the relative abundance among the SMP samples. The binomial (presence/absence) *P* values of the host enrichment analysis of “*Ca. Synechococcus spongiarum*” among various sponge species and environmental samples were calculated as the binomial cumulative probability of the absence of OTUs as described previously (6), excluding *P* values that were corrected for multiple testing using the `p.adjust()` R function using the false-discovery rate (FDR) correction. Samples with corrected *P* values of <0.1 were considered significantly enriched with “*Ca. Synechococcus spongiarum*.” Only sponge samples that contained the ectosome part and which were not collected from the dark caves were further used to test the enrichment of “*Ca. Synechococcus spongiarum*.”

The PAM fluorometry measurements for each sponge species were obtained from Steindler (38).

SNP calling. `SSRG.pl` (<https://github.com/PombertLab/SNPs/blob/master/SSRG>) was used to create a synthetic library of artificial reads from the genomes of “*Ca. Synechococcus feldmannii*.” Further mapping of the either synthetic or sequenced raw Illumina reads against the reference genomes was done accordingly: (i) `get_SNPs_IB.pl` (<https://github.com/PombertLab/SNPs/blob/master/SSRG>) and `FreeBayes` (140) were used to call the variants between the reads and the reference genome, and (ii) SNPs were annotated using `SNPdat_v1.0.5.pl` (<https://github.com/agdorán/snpdat>).

Data availability. Fourteen sequences of genes contained FN3 domains derived from the genomes of “*Ca. Synechococcus feldmannii*” were deposited in the NCBI GenBank database under accession no. [MK422179](https://doi.org/10.1093/nar/nkz422) to [MK422192](https://doi.org/10.1093/nar/nkz422192). The draft genomes of the *Ca. S. feldmannii* 277cV, 277cI, and 288cV have been deposited under Biosample no. [SAMN10755897](https://doi.org/10.1093/bioinformatics/btq107), [SAMN10755956](https://doi.org/10.1093/bioinformatics/btq1075956), and [SAMN10755957](https://doi.org/10.1093/bioinformatics/btq1075957), respectively (Bioproject no. [PRJNA515489](https://doi.org/10.1093/bioinformatics/btq1075489)).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mSystems.00057-19>.

TEXT S1, PDF file, 0.1 MB.

FIG S1, PDF file, 0.5 MB.

FIG S2, PDF file, 0.4 MB.

FIG S3, PDF file, 2.6 MB.

DATA SET S1, XLSX file, 0.1 MB.

ACKNOWLEDGMENTS

Support for this study was provided by the Israel Science Foundation (grant no. 1243/16: “Identification of Molecular Mechanisms Underlying Sponge-Microbiome Symbiosis”). I.B. was supported by an EMBO Short-Term Fellowship.

Jochen Blom is thanked for assistance with the phylogenomic analysis, run on EDGAR. We thank Maya Britstein for help with collection of sponge samples and Stefan Green, Director of the DNA Services Facility at the University of Illinois at Chicago (UIC), for useful comments and suggestions when planning the sequencing for this project.

REFERENCES

- Li CW, Chen JY, Hua TE. 1998. Precambrian sponges with cellular structures. *Science* 279:879–882. <https://doi.org/10.1126/science.279.5352.879>.
- Botha A, Herd DA, Young I, Allison SG. 2012. The first animals: ca. 760-million-year-old sponge-like fossils from Namibia. *S Afr J Sci* 108: 83–90. <https://doi.org/10.4102/sajs.v108i1.2.658>.
- Kaluzhnaya OV, Kulakova NV, Itskovich VB. 2012. Diversity of polyketide synthase (PKS) genes in metagenomic community of freshwater sponge *Lubomirskia baicalensis*. *Mol Biol* 46:790–795. <https://doi.org/10.1134/S002689331206009X>.
- Taylor MW, Radax R, Steger D, Wagner M. 2007. Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol Mol Biol Rev* 71:295–347. <https://doi.org/10.1128/MMBR.00040-06>.
- Webster NS, Thomas T. 2016. The sponge hologenome. *mBio* 7:e00135–16. <https://doi.org/10.1128/mBio.00135-16>.
- Moitinho-Silva L, Nielsen S, Amir A, Gonzalez A, Ackermann GL, Cerrano C, Astudillo-Garcia C, Easson C, Sipkema D, Liu F, Steinert G, Kotoulas G, McCormack GP, Feng G, Bell JJ, Vicente J, Björk JR, Montoya JM, Olson JB, Reveillaud J, Steindler L, Pineda MC, Marra MV, Ilan M, Taylor MW, Polymenakou P, Erwin PM, Schupp PJ, Simister RL, Knight R, Thacker RW, Costa R, Hill RT, Lopez-Legentil S, Dailianis T, Ravasi T, Hentschel U, Li Z, Webster NS, Thomas T. 2017. The sponge microbiome project. *Gigascience* 6:gix077. <https://doi.org/10.1093/gigascience/gix077>.
- Kopp RE, Kirschvink JL, Hilburn IA, Nash CZ. 2005. The Paleoproterozoic snowball Earth: a climate disaster triggered by the evolution of oxygenic photosynthesis. *Proc Natl Acad Sci U S A* 102:11131–11136. <https://doi.org/10.1073/pnas.0504878102>.
- Holland ND. 2006. The oxygenation of the atmosphere and oceans. *Philos Trans R Soc London B, Biol Sci* 361:903–915.
- Bekker A, Holland HD, Wang PL, Rumble D, Stein HJ, Hannah JL, Coetzee LL, Beukes NJ. 2004. Dating the rise of atmospheric oxygen. *Nature* 427:117–120. <https://doi.org/10.1038/nature02260>.
- Larsson J, Nylander JA, Bergman B. 2011. Genome fluctuations in cyanobacteria reflect evolutionary, developmental and adaptive traits. *BMC Evol Biol* 11:187. <https://doi.org/10.1186/1471-2148-11-187>.
- Sánchez-Baracaldo P. 2015. Origin of marine planktonic cyanobacteria. *Sci Rep* 5:17418. <https://doi.org/10.1038/srep17418>.
- Walter JM, Coutinho FH, Dutilh BE, Swings J, Thompson FL, Thompson CC. 2017. Ecogenomics and taxonomy of Cyanobacteria phylum. *Front Microbiol* 8:2132. <https://doi.org/10.3389/fmicb.2017.02132>.
- Coutinho F, Tschoeke DA, Thompson F, Thompson C. 2016. Comparative genomics of *Synechococcus* and proposal of the new genus *Parasynochococcus*. *PeerJ* 4:e1522. <https://doi.org/10.7717/peerj.1522>.
- Coutinho FH, Dutilh BE, Thompson CC, Thompson FL. 2016. Proposal of fifteen new species of *Parasynochococcus* based on genomic, physiological and ecological features. *Arch Microbiol* 198:973–986. <https://doi.org/10.1007/s00203-016-1256-y>.
- Müller W. 1998. Origin of Metazoa: sponges as living fossils. *Naturwissenschaften* 85:11–25. <https://doi.org/10.1007/s001140050444>.
- Erwin DH, Laflamme M, Tweedt SM, Sperling EA, Pisani D, Peterson KJ. 2011. The Cambrian conundrum: early divergence and later ecological success in the early history of animals. *Science* 334:1091–1097. <https://doi.org/10.1126/science.1206375>.
- Fike DA, Grotzinger JP, Pratt LM, Summons RE. 2006. Oxidation of the Ediacaran Ocean. *Nature* 444:744–747. <https://doi.org/10.1038/nature05345>.
- Sahoo SK, Planavsky NJ, Kendall B, Wang X, Shi X, Scott C, Anbar AD, Lyons TW, Jiang G. 2012. Ocean oxygenation in the wake of the Marinoan glaciation. *Nature* 489:546–549. <https://doi.org/10.1038/nature11445>.
- Canfield DE, Poulton SW, Narbonne GM. 2007. Late-Neoproterozoic deep-ocean oxygenation and the rise of animal life. *Science* 315:92–94. <https://doi.org/10.1126/science.1135013>.
- Och LM, Shields-Zhou GA. 2012. The Neoproterozoic oxygenation event: environmental perturbations and biogeochemical cycling. *Earth-Sci Rev* 110:26–57. <https://doi.org/10.1016/j.earscirev.2011.09.004>.
- Steindler L, Huchon D, Avni A, Ilan M. 2005. 16S rRNA phylogeny of sponge-associated cyanobacteria. *Appl Environ Microbiol* 71: 4127–4131. <https://doi.org/10.1128/AEM.71.7.4127-4131.2005>.
- Simister RL, Deines P, Botté ES, Webster NS, Taylor MW. 2012. Sponge-associated clusters revisited: a comprehensive phylogeny of sponge-associated microorganisms. *Environ Microbiol* 14:517–524. <https://doi.org/10.1111/j.1462-2920.2011.02664.x>.
- Erwin PM, Thacker RW. 2008. Cryptic diversity of the symbiotic cyanobacterium *Synechococcus spongium* among sponge hosts. *Mol Ecol* 17:2937–2947. <https://doi.org/10.1111/j.1365-294X.2008.03808.x>.
- Feldmann J. 1933. Sur quelques cyanophycées vivant dans le tissu des éponges de banyules. *Arch Zool Exp Gén* 75:331–404.
- Usher KM, Fromont J, Sutton DC, Toze S. 2004. The biogeography and phylogeny of unicellular cyanobacterial symbionts in sponges from Australia and the Mediterranean. *Microb Ecol* 48:167–177. <https://doi.org/10.1007/s00248-003-1062-3>.
- 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. <https://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. <https://doi.org/10.1007/s00248-012-0047-5>.
- Oren M, Steindler L, Ilan M. 2005. Transmission, plasticity and the molecular identification of cyanobacterial symbionts in the Red Sea sponge *Diacarnus erythraenus*. *Mar Biol* 148:35–41. <https://doi.org/10.1007/s00227-005-0064-8>.
- Lee OO, Chui PY, Wong YH, Pawlik JR, Qian P-Y. 2009. Evidence for vertical transmission of bacterial symbionts from adult to embryo in the Caribbean sponge *Svenzea zeai*. *Appl Environ Microbiol* 75: 6147–6156. <https://doi.org/10.1128/AEM.00023-09>.
- Webster NS, Taylor MW, Behnam F, Lucker 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. <https://doi.org/10.1111/j.1462-2920.2009.02065.x>.
- Gloekner V, Lindquist N, Schmitt S, Hentschel U. 2013. Ectyoplasia ferox, an experimentally tractable model for vertical microbial transmission in marine sponges. *Microb Ecol* 65:462–474. <https://doi.org/10.1007/s00248-012-0142-7>.
- Maldonado M, Riesgo A. 2009. Gametogenesis, embryogenesis, and larval features of the oviparous sponge *Petrosia ficiformis* (Haplosclerida, Demospongiae). *Mar Biol* 156:2181–2197. <https://doi.org/10.1007/s00227-009-1248-4>.
- Lepore E, Sciscioli M, Gherardi M, Scalera-Liaci L. 1995. The ultrastructure of the mature oocyte and the nurse cells of the ceractinomorph *Petrosia ficiformis*. *Cah Biol Mar* 36:15–20.
- Moran NA, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* 42:165–190. <https://doi.org/10.1146/annurev.genet.41.110306.130119>.
- Otero-Bravo A, Goffredi S, Sabree ZL. 2018. Cladogenesis and genomic streamlining in extracellular endosymbionts of tropical stink bugs. *Genome Biol Evol* 10:680–693. <https://doi.org/10.1093/gbe/evy033>.
- 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 spongium*.” *mBio* 5:e00079-14. <https://doi.org/10.1128/mBio.00079-14>.
- 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:e00391-15. <https://doi.org/10.1128/mBio.00391-15>.
- Steindler L. 2000. Energetic and photoadaptational aspects of the association between photosymbionts and marine sponges. Tel Aviv University, Tel Aviv, Israel.
- Marin B, Nowack ECM, Glöckner G, Melkonian M. 2007. The ancestor of the *Paulinella* chromatophore obtained a carboxysomal operon by horizontal gene transfer from a *Nitrococcus*-like gamma-proteobacterium. *BMC Evol Biol* 7:85. <https://doi.org/10.1186/1471-2148-7-85>.
- Makarova KS, Wolf YI, Snir S, Koonin EV. 2011. Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J Bacteriol* 193:6039–6056. <https://doi.org/10.1128/JB.05535-11>.
- Hughes KT, Mathee K. 1998. The anti-sigma factors. *Annu Rev Microbiol* 52:231–286. <https://doi.org/10.1146/annurev.micro.52.1.231>.

42. Aiba H, Mizuno T. 1990. Phosphorylation of a bacterial activator protein, ompR, by a protein kinase, EnvZ, stimulates the transcription of the ompF and ompC genes in *Escherichia coli*. *FEBS Lett* 261:19–22. [https://doi.org/10.1016/0014-5793\(90\)80626-T](https://doi.org/10.1016/0014-5793(90)80626-T).
43. Usher KM, Bergman B, Raven JA. 2007. Exploring cyanobacterial mutualisms. *Annu Rev Ecol Evol Syst* 38:255–273. <https://doi.org/10.1146/annurev.ecolsys.38.091206.095641>.
44. Snyder DS, Brahmasha B, Azadi P, Palenik B. 2009. Structure of compositionally simple lipopolysaccharide from marine *Synechococcus*. *J Bacteriol* 191:5499–5509. <https://doi.org/10.1128/JB.00121-09>.
45. Simkovsky R, Daniels EF, Tang K, Huynh SC, Golden SS, Brahmasha 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. <https://doi.org/10.1073/pnas.1214904109>.
46. Shcolnick S, Keren N. 2006. Metal homeostasis in cyanobacteria and chloroplasts. Balancing benefits and risks to the photosynthetic apparatus. *Plant Physiol* 141:805–810. <https://doi.org/10.1104/pp.106.079251>.
47. Hopkinson BM, Morel F. 2009. The role of siderophores in iron acquisition by photosynthetic marine microorganisms. *Biometals* 22: 659–669. <https://doi.org/10.1007/s10534-009-9235-2>.
48. Durham KA, Porta D, McKay RML, Bullerjahn GS. 2003. Expression of the iron-responsive irpA gene from the cyanobacterium *Synechococcus* sp. strain PCC 7942. *Arch Microbiol* 179:131–134. <https://doi.org/10.1007/s00203-002-0503-6>.
49. Nodop A, Pietsch D, Hocker R, Becker A, Pistorius EK, Forchhammer K, Michel K-P. 2008. Transcript profiling reveals new insights into the acclimation of the mesophilic fresh-water cyanobacterium *Synechococcus elongatus* PCC 7942 to iron starvation. *Plant Physiol* 147: 747–763. <https://doi.org/10.1104/pp.107.114058>.
50. Reddy KJ, Bullerjahn GS, Sherman DM, Sherman LA. 1988. Cloning, nucleotide sequence, and mutagenesis of a gene (irpA) involved in iron-deficient growth of the cyanobacterium *Synechococcus* sp. strain PCC7942. *J Bacteriol* 170:4466–4476. <https://doi.org/10.1128/jb.170.10.4466-4476.1988>.
51. Voth DE. 2011. THANKs for the repeat intracellular pathogens exploit a common eukaryotic domain. *Cell Logist* 1:128–132. <https://doi.org/10.4161/cl.1.4.18738>.
52. Pan X, Lührmann A, Satoh A, Laskowski-Arce MA, Roy CR. 2008. Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. *Science* 320:1651–1654. <https://doi.org/10.1126/science.1158160>.
53. Iturbe-Ormaetxe I, Burke GR, Riegler M, O'Neill SL. 2005. Distribution, expression, and motif variability of ankyrin domain genes in *Wolbachia pipientis*. *J Bacteriol* 187:5136–5145. <https://doi.org/10.1128/JB.187.15.5136-5145.2005>.
54. Kampfraath AA, Klasson L, Anvar SY, Vossen R, Roelofs D, Kraaijeveld K, Ellers J. 2019. Genome expansion of an obligate parthenogenesis-associated *Wolbachia* poses an exception to the symbiont reduction model. *BMC Genomics* 20:106. <https://doi.org/10.1186/s12864-019-5492-9>.
55. Flanagan RC, Neal-McKinney JM, Dhillon AS, Miller WG, Konkel ME. 2009. Examination of *Campylobacter jejuni* putative adhesins leads to the identification of a new protein, designated FlpA, required for chicken colonization. *Infect Immun* 77:2399–2407. <https://doi.org/10.1128/IAI.01266-08>.
56. Johnson BR, Hymes J, Sanozky-Dawes R, Henriksen EDC, Barrangou R, Klaenhammer TR. 2016. Conserved S-layer-associated proteins revealed by exoproteomic survey of S-layer-forming lactobacilli. *Appl Environ Microbiol* 82:134–145. <https://doi.org/10.1128/AEM.01968-15>.
57. Broderick TJ, Duong T. 2016. Mechanisms of *Lactobacillus* persistence and colonization in the gastrointestinal tract of poultry, a review. *Int J Probiotics Prebiotics* 11:15–28.
58. Hymes JP, Johnson BR, Barrangou R, Klaenhammer TR. 2016. Functional analysis of an S-layer-associated fibronectin-binding protein in *Lactobacillus acidophilus* NCFM. *Appl Environ Microbiol* 82:2676–2685. <https://doi.org/10.1128/AEM.00024-16>.
59. 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. <https://doi.org/10.1038/ismej.2010.74>.
60. 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. <https://doi.org/10.1073/pnas.1203287109>.
61. Liu M, Fan L, Zhong L, Kjelleberg S, Thomas T. 2012. Metaproteomic analysis of a community of sponge symbionts. *ISME J* 6:1515–1525. <https://doi.org/10.1038/ismej.2012.1>.
62. Liu MY, Kjelleberg S, Thomas T. 2011. Functional genomic analysis of an uncultured δ -proteobacterium in the sponge *Cymbastela concentrica*. *ISME J* 5:427–435. <https://doi.org/10.1038/ismej.2010.139>.
63. Kamke J, Rinke C, Schwientek P, Mavromatis K, Ivanova N, Sczyrba A, Woyke T, Hentschel U. 2014. The candidate phylum Poribacteria by single-cell genomics: new insights into phylogeny, cell-compartmentalization, eukaryote-like repeat proteins, and other genomic features. *PLoS One* 9:e87353. <https://doi.org/10.1371/journal.pone.0087353>.
64. Diez-Vives C, Moitinho-Silva L, Nielsen S, Reynolds D, Thomas T. 2017. Expression of eukaryotic-like protein in the microbiome of sponges. *Mol Ecol* 26:1432–1451. <https://doi.org/10.1111/mec.14003>.
65. Nguyen M, Liu M, Thomas T. 2014. Ankyrin-repeat proteins from sponge symbionts modulate amoebal phagocytosis. *Mol Ecol* 23: 1635–1645. <https://doi.org/10.1111/mec.12384>.
66. Reynolds D, Thomas T. 2016. Evolution and function of eukaryotic-like proteins from sponge symbionts. *Mol Ecol* 25:5242–5253. <https://doi.org/10.1111/mec.13812>.
67. Hentschel U, Piel J, Degnan SM, Taylor MW. 2012. Genomic insights into the marine sponge microbiome. *Nat Rev Microbiol* 10:641–654. <https://doi.org/10.1038/nrmicro2839>.
68. Brook S. 1998. Pathogenic adaptation of *Escherichia coli* by natural variation of the FimH adhesin. *Proc Natl Acad Sci U S A* 95:8922–8926.
69. Palenik B, Ren Q, Dupont CL, Myers GS, Heidelberg JF, Badger JH, Madupu R, Nelson WC, Brinkac LM, Dodson RJ, Durkin AS, Daugherty SC, Sullivan SA, Khouri H, Mohamoud Y, Halpin R, Paulsen IT. 2006. Genome sequence of *Synechococcus* CC9311: insights into adaptation to a coastal environment. *Proc Natl Acad Sci U S A* 103:13555–13559. <https://doi.org/10.1073/pnas.0602963103>.
70. Avrani S, Wurtzel O, Sharon I, Sorek R, Lindell D. 2011. Genomic island variability facilitates *Prochlorococcus*-virus coexistence. *Nature* 474: 604–608. <https://doi.org/10.1038/nature10172>.
71. Duffresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik BP, Paulsen IT, de Marsac NT, Wincker P, Dossat C, Ferreira 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. <https://doi.org/10.1186/gb-2008-9-5-r90>.
72. Rodriguez-Valera F, Martin-Cuadrado A-B, Rodriguez-Brito B, Pasić L, Thingstad TF, Rohwer F, Mira A. 2009. Explaining microbial population genomics through phage predation. *Nat Rev Microbiol* 7:828–836. <https://doi.org/10.1038/nrmicro2235>.
73. Beiko RG, Harlow TJ, Ragan MA. 2005. Highways of gene sharing in prokaryotes. *Proc Natl Acad Sci U S A* 102:14332–14337. <https://doi.org/10.1073/pnas.0504068102>.
74. Zhaxybayeva O, Gogarten JP, Charlebois RL, Doolittle WF, Papke RT. 2006. Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. *Genome Res* 16:1099–1108. <https://doi.org/10.1101/gr.5322306>.
75. Palenik B, Ren Q, Tai V, Paulsen IT. 2009. Coastal *Synechococcus* metagenome reveals major roles for horizontal gene transfer and plasmids in population diversity. *Environ Microbiol* 11:349–359. <https://doi.org/10.1111/j.1462-2920.2008.01772.x>.
76. Horn H, Slaby B, Jahn M, Bayer K, Moitinho-Silva L, Förster F, Abdelmohsen UR, Hentschel U. 2016. An enrichment of CRISPR and other defense-related features in marine sponge-associated microbial metagenomes. *Front Microbiol* 7:1751. <https://doi.org/10.3389/fmicb.2016.01751>.
77. Slaby BM, Hackl T, Horn H, Bayer K, Hentschel U. 2017. Metagenomic binning of a marine sponge microbiome reveals unity in defense but metabolic specialization. *ISME J* 11:2465–2478. <https://doi.org/10.1038/ismej.2017.101>.
78. Karimi E, Slaby BM, Soares AR, Blom J, Hentschel U, Costa R. 2018. Metagenomic binning reveals versatile nutrient cycling and distinct adaptive features in alphaproteobacterial symbionts of marine sponges. *FEMS Microbiol Ecol*. <https://doi.org/10.1093/femsec/fy074>.
79. Astudillo-García C, Slaby BM, Waite DW, Bayer K, Hentschel U, Taylor MW. 2018. Phylogeny and genomics of SAUL, an enigmatic bacterial lineage frequently associated with marine sponges. *Environ Microbiol* 20:561–576. <https://doi.org/10.1111/1462-2920.13965>.
80. Vasu K, Nagaraja V. 2013. Diverse functions of restriction-modification

- systems in addition to cellular defense. *Microbiol Mol Biol Rev* 77: 53–72. <https://doi.org/10.1128/MMBR.00044-12>.
81. Murray NE. 2000. Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol Mol Biol Rev* 64:412–434. <https://doi.org/10.1128/MMBR.64.2.412-434.2000>.
 82. Waldron DE, Lindsay JA. 2006. SauI: a novel lineage-specific type I restriction-modification system that blocks horizontal gene transfer into *Staphylococcus aureus* and between *S. aureus* isolates of different lineages. *J Bacteriol* 188:5578–5585. <https://doi.org/10.1128/JB.00418-06>.
 83. Levin BR, Antonovics J, Sharma H. 1988. Frequency-dependent selection in bacterial populations. *Philos Trans R Soc Lond B Biol Sci* 319: 459–472. <https://doi.org/10.1098/rstb.1988.0059>.
 84. Smith GR. 2012. How RecBCD enzyme and Chi promote DNA break repair and recombination: a molecular biologist's view. *Microbiol Mol Biol Rev* 76:217–228. <https://doi.org/10.1128/MMBR.05026-11>.
 85. Spies M, Kowalczykowski SC. 2005. Homologous recombination by the RecBCD and RecF pathways, p 389–404. In Higgins NP (ed), *The bacterial chromosome*. ASM Press, Washington, DC.
 86. Bailly X, Olivieri I, Brunel B, Cleyet-Marel J-C, Bena G. 2007. Horizontal gene transfer and homologous recombination drive the evolution of the nitrogen-fixing symbionts of *Medicago* species. *J Bacteriol* 189: 5223–5236. <https://doi.org/10.1128/JB.00105-07>.
 87. Oliveira PH, Touchon M, Cury J, Rocha E. 2017. The chromosomal organization of horizontal gene transfer in bacteria. *Nat Commun* 8:841. <https://doi.org/10.1038/s41467-017-00808-w>.
 88. Krzywinski E, Krzywinski J, Schorey JS. 2004. Naturally occurring horizontal gene transfer and homologous recombination in *Mycobacterium*. *Microbiology* 150:1707–1712. <https://doi.org/10.1099/mic.0.27088-0>.
 89. Price MN, Dehal PS, Arkin AP. 2008. Horizontal gene transfer and the evolution of transcriptional regulation in *Escherichia coli*. *Genome Biol* 9:R4. <https://doi.org/10.1186/gb-2008-9-1-r4>.
 90. Repar J, Briški N, Buljubašić M, Zahradka K, Zahradka D. 2013. Exonuclease VII is involved in “reckless” DNA degradation in UV-irradiated *Escherichia coli*. *Mutat Res* 750:96–104. <https://doi.org/10.1016/j.mrgentox.2012.10.005>.
 91. Chase JW, Richardson CC. 1974. Exonuclease VII of *Escherichia coli*. *J Biol Chem* 249:4553–4560.
 92. Soumitra KS, Tomomori C, Rieko I, Liu D, Tong KI, Park H, Dutta R, Qin L, Swindells MB, Yamazaki T, Ono AM, Kainosho M, Inouye M, Ikura M. 1998. NMR structure of the histidine kinase domain of the *E. coli* osmosensor EnvZ. *Nature* 396:88–92. <https://doi.org/10.1038/23968>.
 93. Meetam 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. <https://doi.org/10.1104/pp.121.4.1267>.
 94. Sveshnikov D, Funk C, Schröder WP. 2007. The PspP-like protein (sl11418) of *Synechocystis* sp. PCC 6803 stabilises the donor side of photosystem II. *Photosynth Res* 93:101–109. <https://doi.org/10.1007/s11120-007-9171-3>.
 95. Jiang H, Kong R, Xu X. 2010. The N-acetylmuramic acid 6-phosphate etherase gene promotes growth and cell differentiation of cyanobacteria under light-limiting conditions. *J Bacteriol* 192:2239–2245. <https://doi.org/10.1128/JB.01661-09>.
 96. Hosoya R, Hamana K, Isobe A, Yokota A. 2005. Polyamine distribution profiles within Cyanobacteria. *Microbiol Cult Collect* 21:3–8.
 97. Hamana K, Furuchi T, Hayashi H, Niitsu M. 2016. Additional analysis of cyanobacterial polyamines—distributions of spermidine, homospermidine, spermine, and thermospermine within the phylum Cyanobacteria. *Microbiol Cult Collect* 32:179–186.
 98. Zhu X, Li Q, Yin C, Fang X, Xu X. 2015. Role of spermidine in overwintering of cyanobacteria. *J Bacteriol* 197:2325–2334. <https://doi.org/10.1128/JB.00153-15>.
 99. Jantaro S, Mäenpää P, Mulo P, Incharoensakdi A. 2003. Content and biosynthesis of polyamines in salt and osmotically stressed cells of *Synechocystis* sp. PCC 6803. *FEMS Microbiol Lett* 228:129–135. [https://doi.org/10.1016/S0378-1097\(03\)00747-X](https://doi.org/10.1016/S0378-1097(03)00747-X).
 100. De Been M, Francke C, Siezen RJ, Abee T. 2011. Novel σ B regulation modules of Gram-positive bacteria involve the use of complex hybrid histidine kinases. *Microbiology* 157:3–12. <https://doi.org/10.1099/mic.0.045740-0>.
 101. Vijay K, Brody MS, Fredlund E, Price CW. 2000. A PP2C phosphatase containing a PAS domain is required to convey signals of energy stress to the σ B transcription factor of *Bacillus subtilis*. *Mol Microbiol* 35: 180–188. <https://doi.org/10.1046/j.1365-2958.2000.01697.x>.
 102. Wilde A, Mullineaux CW. 2015. Motility in cyanobacteria: polysaccharide tracks and type IV pilus motors. *Mol Microbiol* 98:998–1001. <https://doi.org/10.1111/mmi.13242>.
 103. Khayatan B, Meeks JC, Risser DD. 2015. Evidence that a modified type IV pilus-like system powers gliding motility and polysaccharide secretion in filamentous cyanobacteria. *Mol Microbiol* 98:1021–1036. <https://doi.org/10.1111/mmi.13205>.
 104. Taguchi F, Ichinose Y. 2011. Role of type IV pili in virulence of *Pseudomonas syringae* pv. tabaci 6605: correlation of motility, multidrug resistance, and HR-inducing activity on a nonhost plant. *Mol Plant Microbe Interact* 24:1001–1011. <https://doi.org/10.1094/MPMI-02-11-0026>.
 105. Mattick JS. 2002. Type IV pili and twitching motility. *Annu Rev Microbiol* 56:289–314. <https://doi.org/10.1146/annurev.micro.56.012302.160938>.
 106. Leong CG, Bloomfield RA, Boyd CA, Dornbusch AJ, Lieber L, Liu F, Owen A, Slay E, Lang KM, Lostroh CP. 2017. The role of core and accessory type IV pilus genes in natural transformation and twitching motility in the bacterium *Acinetobacter baylyi*. *PLoS One* 12:e0182139. <https://doi.org/10.1371/journal.pone.0182139>.
 107. Gurgui C, Piel J. 2010. Metagenomic approaches to identify and isolate bioactive natural products from microbiota of marine sponges. *Metagenomics Methods Protoc* 668:247–264. https://doi.org/10.1007/978-1-60761-823-2_17.
 108. Joshi NA, Fass JN. 2011. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files. <https://github.com/najoshi/sickle>.
 109. Andrews S. 2014. FastQC a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
 110. Peng Y, Leung HCM, Yiu SM, Chin F. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28:1420–1428. <https://doi.org/10.1093/bioinformatics/bts174>.
 111. 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. <https://doi.org/10.1038/ismej.2012.148>.
 112. Laczny CC, Sternal T, Plugaru V, Gawron P, Atashpendar A, Margossian H, Coronado S, der Maaten L, Vlassis N, Wilmes P. 2015. VizBin—an application for reference-independent visualization and human-augmented binning of metagenomic data. *Microbiome* 3:1. <https://doi.org/10.1186/s40168-014-0066-1>.
 113. Hunt M, Kikuchi T, Sanders M, Newbold C, Berriman M, Otto TD. 2013. REAPR: a universal tool for genome assembly evaluation. *Genome Biol* 14:R47. <https://doi.org/10.1186/gb-2013-14-5-r47>.
 114. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>.
 115. 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. <https://doi.org/10.1186/1471-2105-11-119>.
 116. Liu W, Xie Y, Ma J, Luo X, Nie P, Zuo Z, Lahrmann U, Zhao Q, Zheng Y, Zhao Y, Xue Y, Ren J. 2015. IBS: an illustrator for the presentation and visualization of biological sequences. *Bioinformatics* 31:3359–3361. <https://doi.org/10.1093/bioinformatics/btv362>.
 117. Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res* 43:D261–D269. <https://doi.org/10.1093/nar/gku1223>.
 118. de Crécy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Rückert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* 33:5691–5702. <https://doi.org/10.1093/nar/gki866>.
 119. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Von Mering C, Bork P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol Biol Evol* 34: 2115–2122. <https://doi.org/10.1093/molbev/msx148>.

120. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, Jensen LJ, Von Mering C, Bork P. 2016. EGGNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* 44:D286–D293. <https://doi.org/10.1093/nar/gkv1248>.
121. Geer LY, Domrachev M, Lipman DJ, Bryant SH. 2002. CDART: protein homology by domain architecture. *Genome Res* 12:1619–1623. <https://doi.org/10.1101/gr.278202>.
122. Markowitz VM, Chen I-M, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC. 2012. IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res* 40:D115–D122. <https://doi.org/10.1093/nar/gkr1044>.
123. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. *Nucleic Acids Res* 47:D427–D432. <https://doi.org/10.1093/nar/gky995>.
124. Faith DP, Minchin PR, Belbin L. 1987. Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* 69:57–68. <https://doi.org/10.1007/BF00038687>.
125. Minchin PR. 1987. An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio* 69:89–107. <https://doi.org/10.1007/BF00038690>.
126. Barter RL, Yu B. 2018. Superheat: an R package for creating beautiful and extendable heatmaps for visualizing complex data. *J Comput Graph Stat* 27:910–922. <https://doi.org/10.1080/10618600.2018.1473780>.
127. 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. <https://doi.org/10.1186/1471-2105-10-154>.
128. 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. <https://doi.org/10.1186/1471-2164-9-75>.
129. 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. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* 42:D206–D214. <https://doi.org/10.1093/nar/gkt1226>.
130. Grissa I, Vergnaud G, Pourcel C. 2007. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 8:172. <https://doi.org/10.1186/1471-2105-8-172>.
131. Sela I, Ashkenazy H, Katoh K, Pupko T. 2015. GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res* 43:W7–W14. <https://doi.org/10.1093/nar/gkv318>.
132. Penn O, Privman E, Ashkenazy H, Landan G, Graur D, Pupko T. 2010. GUIDANCE: a web server for assessing alignment confidence scores. *Nucleic Acids Res* 38(Web Server issue):W23–W28. <https://doi.org/10.1093/nar/gkq443>.
133. Landan G, Graur D. 2008. Local reliability measures from sets of co-optimal multiple sequence alignments. *Pac Symp Biocomput* 2008: 15–24.
134. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>.
135. Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Mol Biol Evol* 25:1307–1320. <https://doi.org/10.1093/molbev/msn067>.
136. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33: 1870–1874. <https://doi.org/10.1093/molbev/msw054>.
137. Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23:127–128. <https://doi.org/10.1093/bioinformatics/btl529>.
138. Thomas T, Moitinho-Silva L, Lurgi M, Björk JR, Easson C, Astudillo-García C, Olson JB, Erwin PM, López-Legentil S, Luter H, Chaves-Fonnegra A, Costa R, Schupp PJ, Steindler L, Erpenbeck D, Gilbert J, Knight R, Ackermann G, Victor Lopez J, Taylor MW, Thacker RW, Montoya JM, Hentschel U, Webster NS. 2016. Diversity, structure and convergent evolution of the global sponge microbiome. *Nat Commun* 7:11870. <https://doi.org/10.1038/ncomms11870>.
139. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
140. Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv* 1207.3907. <https://arxiv.org/abs/1207.3907>.
141. Díez-Vives C, Esteves AIS, Costa R, Nielsen S, Thomas T. 2018. Detecting signatures of a sponge-associated lifestyle in bacterial genomes. *Environ Microbiol Rep* 10:433–443. <https://doi.org/10.1111/1758-2229.12655>.