




# Comparative Genomics of Cyanobacterial Symbionts Reveals Distinct, Specialized Metabolism in Tropical *Dysideidae* Sponges

Michelle A. Schorn,<sup>a</sup> Peter A. Jordan,<sup>a</sup> Sheila Podell,<sup>a</sup> Jessica M. Blanton,<sup>a</sup> Vinayak Agarwal,<sup>a,b</sup> Jason S. Biggs,<sup>c</sup> Eric E. Allen,<sup>a,d,e</sup>  Bradley S. Moore<sup>a,d,f</sup>

<sup>a</sup>Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California, San Diego, California, USA

<sup>b</sup>School of Chemistry and Biochemistry, School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia, USA

<sup>c</sup>University of Guam Marine Laboratory, UoG Station, Mangilao, Guam, USA

<sup>d</sup>Center for Microbiome Innovation, University of California, San Diego, California, USA

<sup>e</sup>Division of Biological Sciences, University of California, San Diego, California, USA

<sup>f</sup>Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, California, USA

**ABSTRACT** Marine sponges are recognized as valuable sources of bioactive metabolites and renowned as petri dishes of the sea, providing specialized niches for many symbiotic microorganisms. Sponges of the family *Dysideidae* are well documented to be chemically talented, often containing high levels of polyhalogenated compounds, terpenoids, peptides, and other classes of bioactive small molecules. This group of tropical sponges hosts a high abundance of an uncultured filamentous cyanobacterium, *Hormoscilla spongelliae*. Here, we report the comparative genomic analyses of two phylogenetically distinct *Hormoscilla* populations, which reveal shared deficiencies in essential pathways, hinting at possible reasons for their uncultivable status, as well as differing biosynthetic machinery for the production of specialized metabolites. One symbiont population contains clustered genes for expanded polybrominated diphenylether (PBDE) biosynthesis, while the other instead harbors a unique gene cluster for the biosynthesis of the dysinosin nonribosomal peptides. The hybrid sequencing and assembly approach utilized here allows, for the first time, a comprehensive look into the genomes of these elusive sponge symbionts.

**IMPORTANCE** Natural products provide the inspiration for most clinical drugs. With the rise in antibiotic resistance, it is imperative to discover new sources of chemical diversity. Bacteria living in symbiosis with marine invertebrates have emerged as an untapped source of natural chemistry. While symbiotic bacteria are often recalcitrant to growth in the lab, advances in metagenomic sequencing and assembly now make it possible to access their genetic blueprint. A cell enrichment procedure, combined with a hybrid sequencing and assembly approach, enabled detailed genomic analysis of uncultivated cyanobacterial symbiont populations in two chemically rich tropical marine sponges. These population genomes reveal a wealth of secondary metabolism potential as well as possible reasons for historical difficulties in their cultivation.

**KEYWORDS** biosynthesis, cyanobacteria, halogenated compounds, marine microbiology, metagenomics, natural products, nonribosomal peptide synthetase

Elements of natural selection have produced exquisite, naturally derived chemical diversity that humans have exploited for medicinal potential for centuries, including the majority of approved pharmaceutical drugs (1). Marine invertebrates, particularly

**Citation** Schorn MA, Jordan PA, Podell S, Blanton JM, Agarwal V, Biggs JS, Allen EE, Moore BS. 2019. Comparative genomics of cyanobacterial symbionts reveals distinct, specialized metabolism in tropical *Dysideidae* sponges. *mBio* 10:e00821-19. <https://doi.org/10.1128/mBio.00821-19>.

**Editor** Margaret J. McFall-Ngai, University of Hawaii at Manoa

**Copyright** © 2019 Schorn 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/).

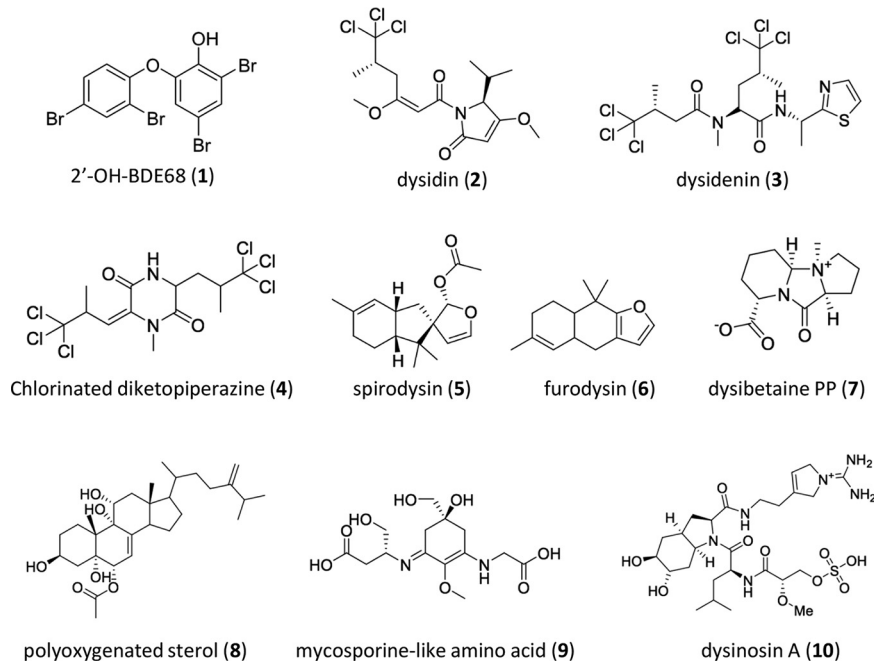
Address correspondence to Bradley S. Moore, [bsmoore@ucsd.edu](mailto:bsmoore@ucsd.edu).

This article is a direct contribution from a Fellow of the American Academy of Microbiology. Solicited external reviewers: Mohamed Donia, Princeton University; Robert Thacker, Stony Brook University.

**Received** 3 April 2019

**Accepted** 9 April 2019

**Published** 14 May 2019



**FIG 1** Representative secondary metabolites previously isolated from *Lamellodysidea herbacea* specimens.

sponges, have emerged as highly rewarding reservoirs of natural chemical diversity (2). These ancient, sessile animals are unable to physically evade predators, and it is thought that many have formed symbioses with microbes that can produce chemical defenses while benefiting from a nutrient-rich environment (3). Access to these symbionts and their full chemical repertoires has been limited due to difficulties in culturing many symbiotic microbes. The genomic revolution has provided new tools for sustainably interrogating complex metagenomes and has exposed the surprising breadth of biosynthetic capabilities of talented natural-product-generating microbes (4).

The *Lamellodysidea* (formerly *Dysidea*) genus within the *Dysideidae* family is particularly prolific in its variety of bioactive natural products (Fig. 1). This family's best-known class of compounds, due to the environmental toxicity of their anthropogenic counterparts, is the polybrominated diphenyl ethers (PBDEs), e.g., compound 1, first isolated in 1972 from *Dysidea herbacea* and subsequently from numerous *Dysideidae* sponges (5, 6). Astonishingly, PBDEs can make up over 10% of the sponge's dry weight, with cell sorting (7) and microbiome sequencing studies (8) attributing these abundant molecules to the dominant cyanobacterial symbiont. Numerous other distinct structural classes of molecules have been isolated from *Lamellodysidea* specimens (Fig. 1), including polychlorinated peptidic molecules (e.g., compounds 2 to 4) (9, 10). Some of these chlorinated metabolites have been credited to the dominant cyanobacterial symbiont rather than the sponge itself (11–13). On the other hand, several distinct sesquiterpene molecules (e.g., compounds 5 and 6) that colocalize with sponge cells rather than cyanobacteria (12) have also been characterized from *Lamellodysidea* (14). At least four other unique classes of metabolites (e.g., compounds 7 to 10) have been isolated from *Lamellodysidea herbacea* with no experimental evidence of the true producer within the sponge holobiont (15–19).

Such a wealth of chemistry from one family of sponges, with a focus on one species here, *Lamellodysidea herbacea*, is remarkable, but with the abundance of microbial biodiversity residing in sponges, the true source of these diverse compounds is difficult to decipher. In some cases, sequencing has shown that the dominant symbiont of an invertebrate assemblage contains the biosynthetic machinery to be a major source of natural products (20). Genomic sequencing has shown this to be the case for the

marine ascidian *Lissoclinum patella* and its uncultured cyanobacterial symbiont, *Prochloron didemni*, and the marine sponge *Theonella swinhoei* and its uncultured bacterial symbiont, *Entotheonella swinhoei* (21, 22).

A defining feature of *Lamellodysidea* sponges is the persistent presence of a filamentous cyanobacterial symbiont, *Hormoscilla spongelliae* (formerly *Oscillatoria spongelliae*) (23), with distinct strains inhabiting morphologically discrete hosts (24, 25). Despite repeated attempts, these symbionts have been recalcitrant to culturing efforts and are therefore assumed to be obligate symbionts, unable to live outside their host system (26). Likewise, field studies have shown that when cyanobacterial photosynthesis is inhibited by shading, *Lamellodysidea* sponges have a higher mortality rate (27), suggesting that the host sponge critically depends on symbiont carbon fixation, making this a bidirectional obligate symbiosis.

Recently, we reported the biosynthetic gene clusters (BGCs) responsible for PBDE synthesis within the genomes of three *Dysideidae* cyanobacterial symbionts, establishing bacterial origins for at least one major metabolite isolated from *L. herbacea* sponges (8). However, origins of the majority of secondary metabolites isolated from *L. herbacea* specimens have not yet been addressed, and the underlying causes of our inability to cultivate many sponge symbionts are still unknown. Obtaining high-quality draft genomes of the uncultured cyanobacterial symbionts has allowed us to investigate both of these questions, providing potential reasons for their seemingly obligate symbiont lifestyle and revealing differing biosynthetic capacities for producing secondary metabolites of varied chemotypes.

## RESULTS AND DISCUSSION

**Enriched MAGs.** Two *L. herbacea* sponges, named GUM007 and GUM202, exhibited different chemotypes: GUM007 contained no PBDEs, while GUM202 contained an abundance of PBDEs. Differences in secondary metabolite biosynthetic potential in the two sponge symbionts were explored using a hybrid sequencing approach to generate metagenome-assembled genomes (MAGs). Short-read metagenomic sequencing alone often fails to produce sufficiently contiguous sequences for examining long, multigene biosynthetic gene clusters (BGCs) (28). Recently developed technologies, such as Pacific Biosciences (PacBio) single-molecule real-time sequencing, enable the high-throughput generation of very long reads. Reduced accuracy of these long reads can be offset by supplementation with high-accuracy, short reads obtained from Illumina sequencing. This strategy was previously shown to improve assembly quality of over 30 binned genomes from a deeply sequenced sponge microbiome (29). To enhance the purity of DNA from yet-uncultivable *Hormoscilla* symbionts, we isolated cyanobacterial trichomes from the less dense sponge cells by pressing the sponge tissue to exude cyanobacterial cells and separating them by centrifugal partitioning (see Fig. S1 in the supplemental material) (26).

Hybrid assemblies using phylogenetically classified Illumina and PacBio reads resulted in high-quality draft MAGs when assessed according to published standard recommendations promulgated by the Genomics Standards Consortium (Table 1) (30). The GUM007\_hs genome consisted of 64 scaffolds, with an average length of 97 kb and 94.46% completeness, according to CheckM (31). GUM202\_hs consisted of 70 scaffolds, with an average length of 98 kb and 97.41% completeness. These two population genomes have an average nucleotide identity (ANI) of 96.18%, an average amino acid identity (AAI) of 93.12%, and a 16S rRNA nucleotide pairwise identity of 97.9%. Recent standards for classifying uncultivated microbes indicate that the evolutionary distance of these two populations is on the threshold between genus and species, with two out of three criteria indicating that they are different species (32). While these two populations are phylogenetically closely related, their overall genome organization contains only limited regions of synteny (Fig. S2). This biological strategy of retaining closely related conserved genes, while allowing greater genomic plasticity in other regions, such as genomic islands where secondary metabolism genes reside, has been reported for other chemically talented bacteria (33).

**TABLE 1** Assembly and quality statistics for GUM\_hs genomes<sup>a</sup>

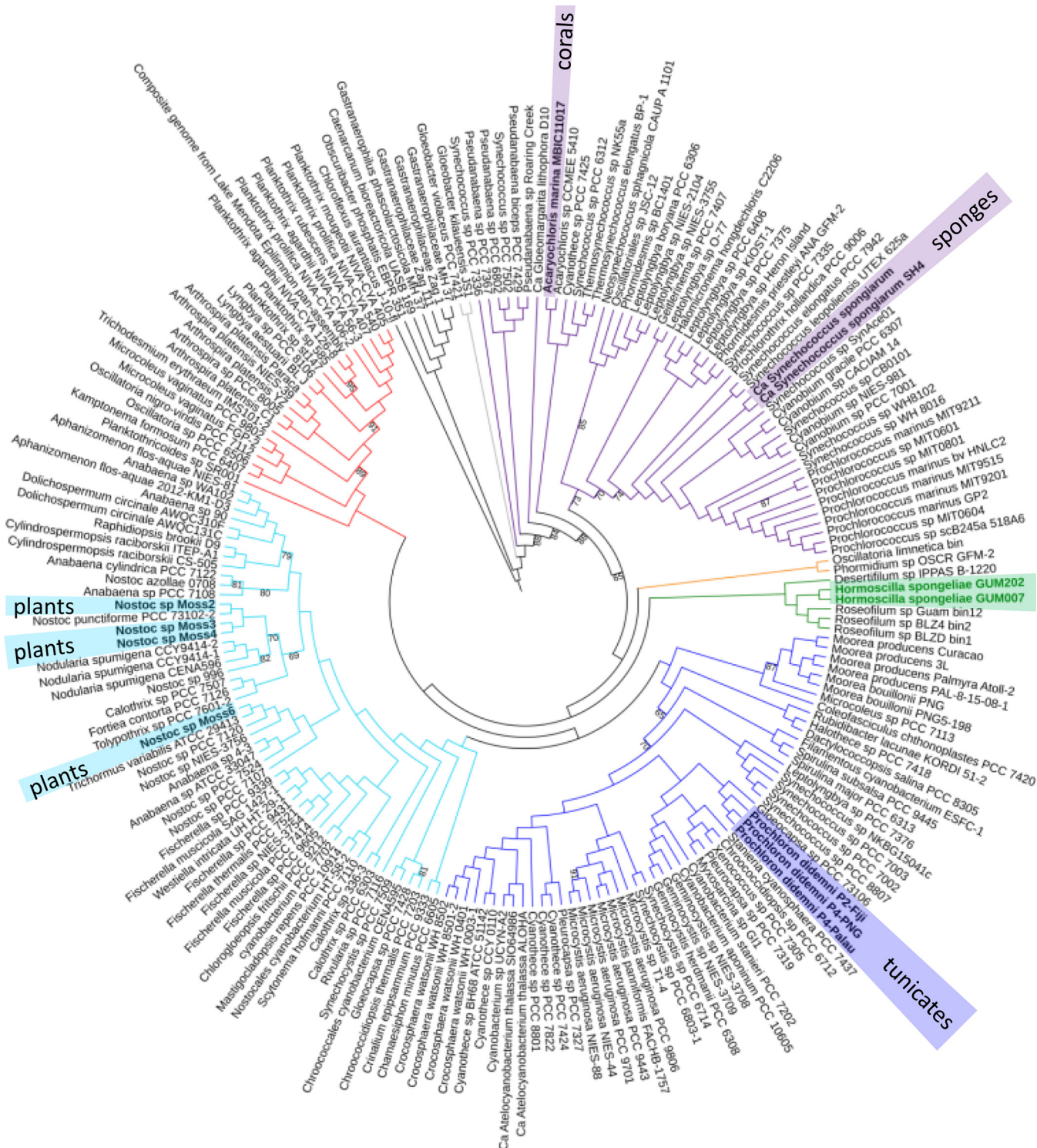
Statistic	Genome	
	GUM007_hs	GUM202_hs
Analysis project type	MAG	MAG
Taxon ID	Multimarker	Multimarker
Assembly software	IDBA-UD, Celera, SSPACE-LongRead	IDBA-UD, hybridSPAdes
Assembly quality	High-quality draft	High-quality draft
rRNA genes	23S, 16S, 5S	23S, 16S, 5S
CheckM completeness (%)	94.46	97.41
CheckM contamination	2.24	2.00
No. of scaffolds	64	70
Avg length of scaffolds (bp)	97,372	98,137
Longest scaffold (bp)	345,613	304,315
Estimated genome size (Mb)	6.2	6.8
$N_{50}$ (bp)	169,688	161,164
% GC	47.8	47.5
Bin parameters	% GC, nucleotide composition, coverage, and taxonomic assignment	% GC, nucleotide composition, coverage, and taxonomic assignment
Binning software	Custom/manual	Custom/manual

<sup>a</sup>General assembly and quality statistics for the *Hormoscilla* metagenome-assembled genomes (MAGs) from GUM007 and GUM202. Both MAGs meet the MIMAG standards for high-quality draft MAGs. Two-way ANI was 96.18%, two-way AAI was 93.12%, and 16S rRNA nucleotide pairwise identity was 97.9%.

A well-supported multilocus sequence analysis phylogenetic tree gives a phylogenetic reference for how these unusual symbionts are related to other cyanobacterial taxa (Fig. 2). The two *Hormoscilla* MAGs clade together and, within the realm of sequenced organisms, are most closely related to multiple *Roseofilum* MAGs. *Roseofilum* spp. are filamentous cyanobacteria found in microbial assemblages that afflict corals with a disease called black band disease (BBD) (34). Comparative genomics of five *Roseofilum* spp., which have never been found living independently, have shown that they are reliant on other members of the microbial consortium, as many filamentous cyanobacteria are, and as *Hormoscilla* appears to be (34). The deep-branching clade that contains both *Hormoscilla* and *Roseofilum* spp. is sparsely populated, perhaps due to their participation in microbial consortia, thus making them difficult to culture.

**Genomic hallmarks of a symbiotic lifestyle.** *Hormoscilla* symbionts were initially expected to contain a pared-down genome with loss of metabolic functions, such as those seen in "*Candidatus* *Synechococcus spongiarum*," yet-uncultured cyanobacterial symbionts of diverse sponges (35, 36). However, estimated sizes of the *Hormoscilla* genomes ranged between 6.2 and 6.8 Mb, comparable to other free-living filamentous cyanobacteria. The *Hormoscilla* genomes did share some other symbiont specific characteristics of "*Ca. Synechococcus spongiarum*" genomes. Ankyrin repeat proteins (COG0666) and leucine-rich repeat (LRR) proteins (COG4886) are present in both *Hormoscilla* genomes and are postulated to play a role in bacterium-sponge recognition. "*Ca. Synechococcus spongiarum*" genomes exhibited a loss of low-molecular-weight peptides involved in photosystem II, and PsbJ was lost in both *Hormoscilla* and PsbX in GUM007\_hs. Both *Hormoscilla* genomes were also missing several oxidative-stress-related genes, including glutathione peroxidase (EC 1.11.1.9), gamma-glutamyl transpeptidase (EC 2.3.2.2), and superoxide dismutase (SOD) (36).

The mounting evidence of uncultured symbionts from marine invertebrates may point to an alternative evolutionary phenomenon: retaining a large, chemically talented genome can impart evolutionary advantages to the holobiont (79). Seemingly obligate endosymbionts such as *Hormoscilla*, *P. didemni* (37), and *Entotheonella* spp. (38), which



**FIG 2** Multilocus sequence analysis of diverse cyanobacteria. One hundred ninety-seven cyanobacterial genomes containing 25 housekeeping genes in single copy were used to construct a maximum-likelihood tree. The two *Hormoscella* spp. are highlighted in green, and their nearest sister clade is the *Roseofilum* spp., known coral pathogens. Other close relatives, *Desertifilum* sp., *Phormidium* sp., and *Oscillatoria limnetica*, have been found growing in microbial mats. Other symbiotic cyanobacteria are highlighted throughout the tree. Only bootstrap values between 65 and 95 are shown; all other bootstrap values are above 95.

retain a large genome despite reliance on their host, may represent another evolutionary mechanism in which the maintenance of extended genomic repertoires sustains the symbiotic interaction. If this were the case, these symbionts might not complete the genome streamlining process because losing their biosynthetic capabilities to manu-

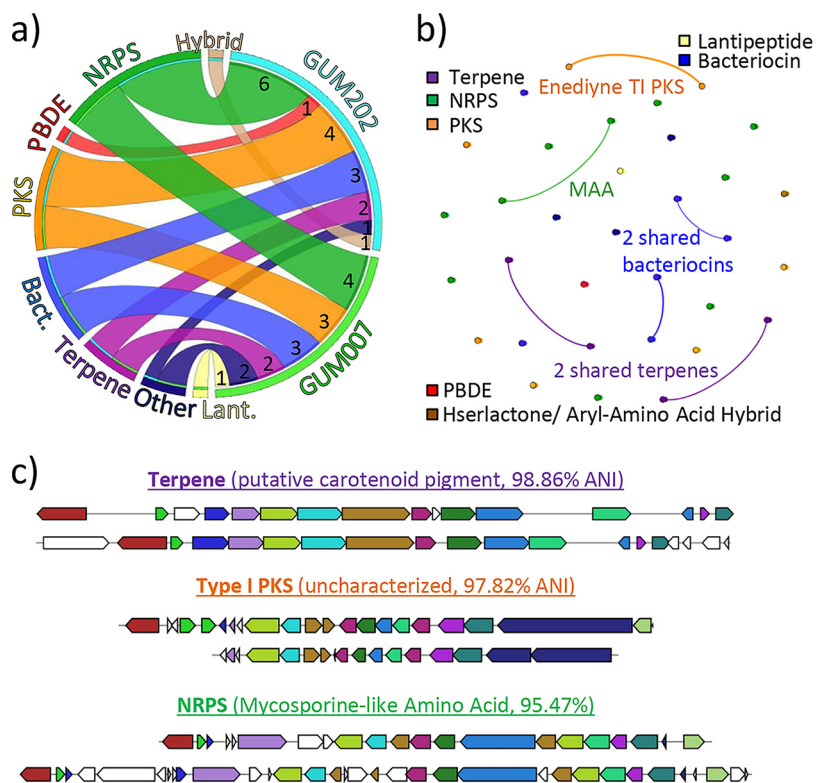
facture metabolically expensive chemicals would reduce the evolutionary fitness of the holobiont.

**Missing essential gene analysis.** To understand why *Hormoscilla* species cannot be cultured in the laboratory despite their lack of canonical symbiotic traits, we compared metabolic pathways and gene essentiality with an extensively characterized free-living cyanobacterium. A curated genome-scale model (GEM) has been developed to determine the genes essential to sustain life in a photosynthetic organism using the model cyanobacterium *Synechococcus elongatus* PCC 7942 (39). This GEM is informed not just by metabolic modeling data but also by genome-wide gene essentiality analysis, determined from ~250,000 transposon mutants to assign genes as essential, beneficial, or nonessential, making it the only large-scale study of the genes essential to sustaining photosynthetic life (40). Although *S. elongatus* and *Hormoscilla* belong to different orders (*Synechococcales* and *Oscillatoriales*, respectively), and differences in gene content are expected, the comparison allowed us to leverage the extensive, experimentally validated data about gene essentiality in a free-living photosynthetic organism to expose missing essential genes in a cyanobacterial symbiont.

An initial comparison of the GUM\_hs genes versus the *S. elongatus* GEM revealed a combined list of 155 genes missing in both GUM\_hs genomes. Missing genes were then converted to corresponding KEGG Ontology (KO) (41) and pfam (42) models for a more robust search. Any models that remained unfound in the *Hormoscilla* MAGs were cross-referenced with gene essentiality data for *S. elongatus* PCC 7942, resulting in seven KOs/pfams that are missing in the GUM\_hs assemblies yet essential for *S. elongatus* (see Table S1 in the supplemental material).

Both population genomes predict that *Hormoscilla* are prototrophic for all amino acids except histidine. The missing essential gene analysis exposed a lack of imidazole glycerol phosphate dehydratase (HisB), the enzyme responsible for the sixth step in histidine biosynthesis, in both GUM\_hs genomes (Fig. S3). No alternative pathway for histidine biosynthesis was found. Due to the phylogenetic distance of *S. elongatus*, a more closely related filamentous cyanobacterium with a complete genome sequence, *Moorea producens*, was also used for comparison (43). A BLAST search using *hisB* from *Moorea producens* did not reveal any candidate orthologs in the GUM\_hs genomes. Furthermore, *hisB* from *M. producens* and the pfam HMM for HisB (PF00475) were used to search the assembled metagenomes, and no significant hits to cyanobacteria were found. While it is possible that this gene was not assembled from the sequence data generated, it may also indicate that *Hormoscilla* species are histidine auxotrophs, requiring histidine supplementation from their hosts or other members of the microbiome. Alternatively, there could be a complementary gene encoding a novel dehydratase not yet discovered to replace HisB in histidine biosynthesis.

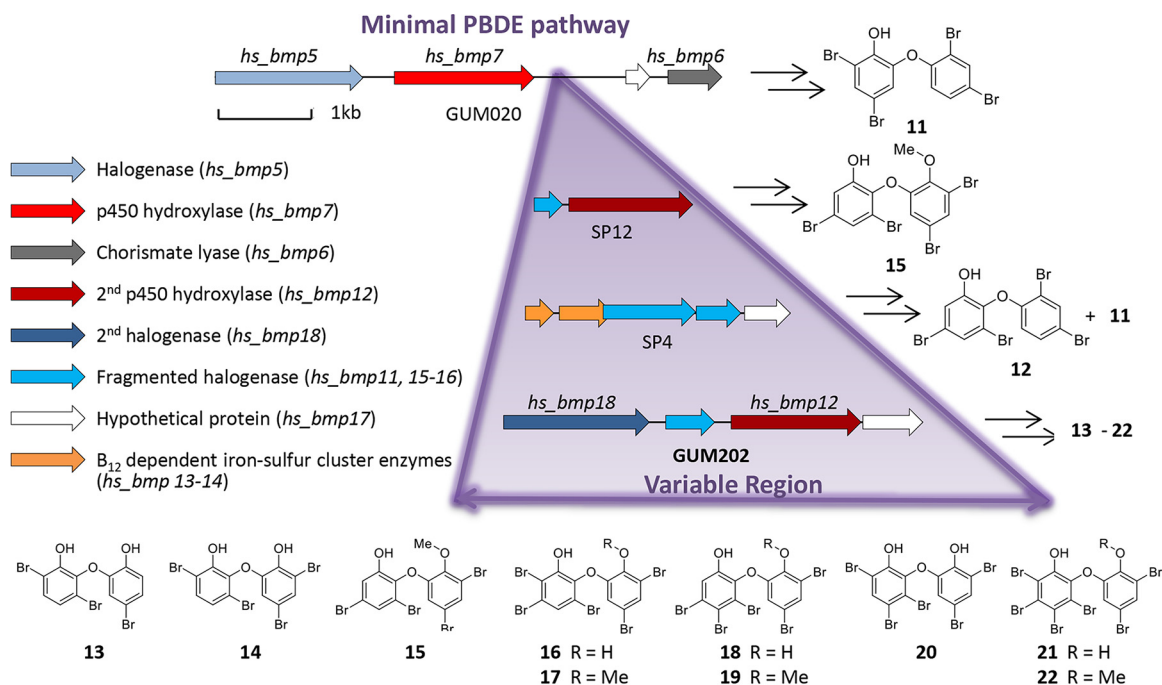
The remaining six genes that are essential for *S. elongatus* but missing in both GUM\_hs assemblies were determined to have complementary pathways for completing the required reactions (Table S1 and Text S1). Additionally, we examined the ability of *Hormoscilla* symbionts to produce important cofactors. Both GUM\_hs genomes are missing canonical genes encoding the enzymes performing the last step in thiamine biosynthesis, which are present in both *S. elongatus* and *M. producens* (Fig. S4a). This suggests that *Hormoscilla* symbionts may be unable to make thiamine. The pathway for production of biotin, an important cofactor in fatty acid and amino acid biosynthesis, is also incomplete in *Hormoscilla*, with key enzymes BioA and BioC missing (Fig. S4b). Interestingly, *M. producens* is also missing BioA, but the entire biotin pathway is present in a persistent, uncultured heterotrophic bacterium growing together with *M. producens* (44). *M. producens* does not grow axenically and always has a consortium of heterotrophic bacteria growing with it; thus, it is possible that neither *M. producens* nor *Hormoscilla* symbionts can make biotin on their own and rely on heterotrophic bacteria for provision of this important cofactor. Future cultivation efforts of *Hormoscilla* symbionts should include addition of histidine, thiamine, and biotin to address possible deficits in these biosynthetic pathways.



**FIG 3** Secondary metabolite biosynthetic gene clusters (BGCs) of two *Hormoscilla* populations. An overview of the variety of BGCs found in the two *Hormoscilla* populations, including shared and distinct BGCs. (a) Number and classes of BGCs in the GUM\_hs genomes. (b) Gene cluster similarity network where each node represents a gene cluster and those with similarity over a 0.4 threshold are connected by a line. The weight of the line indicates higher similarity for bolder lines. (c) MultiGeneBlast comparison of similar clusters displays gene synteny.

**Distinct specialized metabolism of the GUM007 and GUM202 *Hormoscilla* symbionts.** The near-completeness of the two genomes afforded us a comprehensive survey of identifiable BGCs within these *Hormoscilla* specimens. AntiSMASH 4.1.0 (45) results for both GUM007\_hs and GUM202\_hs are shown in Fig. 3a. A total of 15 BGCs (after splitting an unlikely nonribosomal peptide synthetase [NRPS]/bacteriocin hybrid cluster) were found in GUM007\_hs, and 18 were found in GUM202\_hs. A BGC similarity network, created using BiG-SCAPE (46), revealed that the genomes share six clusters, making two-thirds of the BGCs unique to each organism (Fig. 3b). Four out of the six shared clusters have ANI values above the whole-genome ANI score (96.18%), which may suggest that these gene clusters are conserved for the benefit of the symbiosis.

Of those clusters shared, half have predicted structures. Two terpene clusters, with ANIs of 90.24% and 98.86%, likely encode carotenoid pigments, due to the presence of lycopen synthases, and are not predicted to be implicated in the production of characteristic sesquiterpenes isolated from *L. herbacea* sponges. Notably, no sesquiterpene biosynthetic machinery was identified within the *Hormoscilla* genomes, suggesting that another member of the sponge holobiont, possibly the sponge itself, is producing the abundant sesquiterpenes often observed in *L. herbacea*. The shared NRPS cluster, with an ANI of 95.47%, has homology to shinorine, a mycosporine-like amino acid (MAA) (47). MAAs are common cyanobacterial metabolites with UV-protective properties that may also contribute to environment amelioration as compatible osmolytes and antioxidants, suggesting a benefit that the cyanobacterium may provide to its host sponge (48). The remaining common clusters are uncharacterized, including two ribosomally synthesized and posttranslationally modified peptides (RiPPs), with ANIs of 96.74% and 97.56%, and an enediayne type I polyketide synthase



**FIG 4** Expansion of *hs\_bmp* in GUM202\_*hs* leads to structurally varied PBDEs. The *hs\_bmp* cluster, minimally made up of a halogenase (*hs\_bmp5*), a p450 hydroxylase (*hs\_bmp7*), and a chorismate lyase (*hs\_bmp6*), has been shown to be responsible for the production of PBDEs. In addition to these core genes, populations of *Hormoscilla* from different clades of *Dysideidae* sponges (specimens SP12, SP4, and GUM202) contain extra genes in a variable region that correspond to the chemistry seen in each sponge.

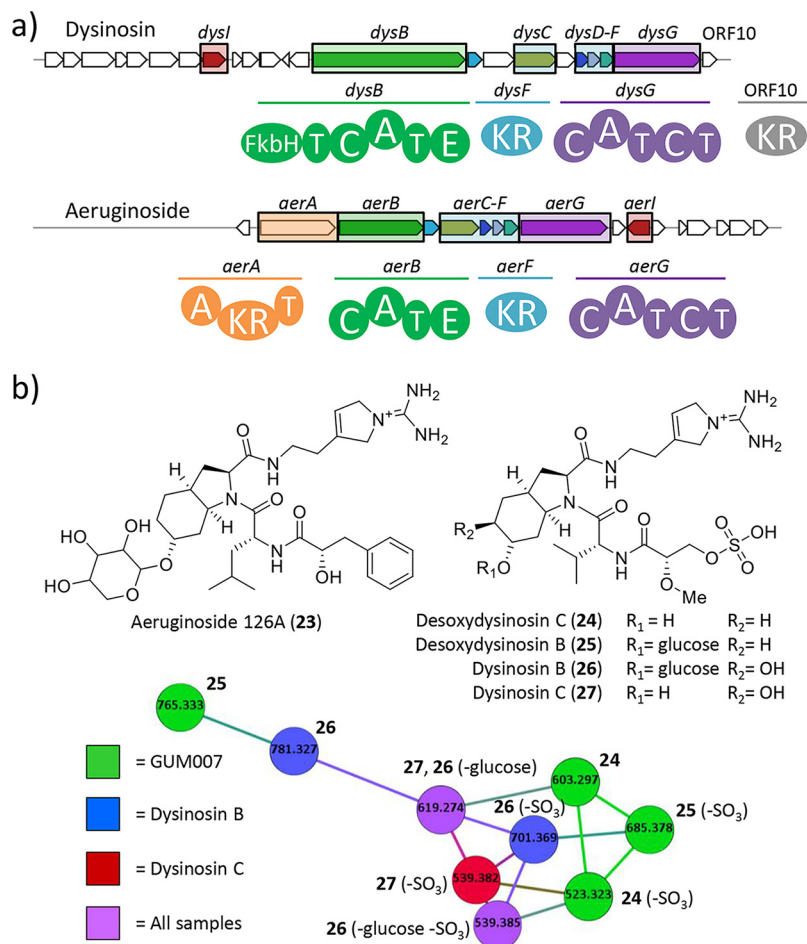
(T1PKS), with an ANI of 97.82%, without predictable product structures. Overall, these filamentous cyanobacteria maintain a large biosynthetic repertoire for making specialized secondary metabolites, as is common in late-branching cyanobacteria (49).

Two of the BGCs, each unique to one of the GUM\_*hs* genomes, were subjected to further chemical characterization. The first is an expanded PBDE cluster in GUM202\_*hs*, larger than any previously described, which contains an extra halogenase gene. GUM007 does not contain PBDEs and lacks the PBDE gene cluster (8) and yet contains a bioinformatically predictable NRPS cluster similar to the aeruginoside BGC (50).

**Genetic expansion in *hs\_bmp* cluster leads to structural variety of PBDEs.** Previous investigations of three *Hormoscilla* symbiont clades revealed that PBDE biosynthesis is encoded in the semivariable *hs\_bmp* gene cluster (Fig. 4) (8). At the minimum, three core enzymes are needed to assemble PBDEs: Bmp5, a flavin-dependent brominase; Bmp6, a chorismate lyase; and Bmp7, a cytochrome P450 that couples the two brominated phenolic rings (51). Using these fundamental biosynthetic features, we queried the genomes of GUM202\_*hs* and GUM007\_*hs*. While the *hs\_bmp* gene cluster is absent from GUM007\_*hs*, in accordance with its secondary metabolite profile, GUM202\_*hs* possesses an expanded *bmp* gene cluster (Fig. 4). Previously, we observed a variable genomic region between *hs\_bmp6* and *hs\_bmp7* in the three sequenced *hs\_bmp* clusters, one of which contains *hs\_bmp12*, a cytochrome P450 hydroxylase that was shown to be responsible for further hydroxylated PBDEs exclusive to clade Ia sponges. The variable region in the GUM202 *hs\_bmp* pathway also contains *hs\_bmp12* with 99.3% pairwise identity to *hs\_bmp12* from clade Ia, as well as a gene encoding a second putative flavin-dependent halogenase, *hs\_bmp18*, with 78% pairwise nucleotide identity to the GUM202 *hs\_bmp5*.

Based on the genomic analysis of the *hs\_bmp* cluster in GUM202\_*hs*, we predicted that PBDEs isolated from this sponge would have a higher degree of halogenation and hydroxylation than PBDEs previously characterized from metagenome-sequenced *L. herbacea* sponges (8). LC-MS analysis of methanol extracts from GUM202 revealed 10 distinct PBDEs (13–22) (Fig. S5), a majority of which are penta- and hexabrominated. We





**FIG 5** Gene cluster comparison and molecular network of dysinosins. Novel desoxydysinosin discovery through genome mining. (a) Comparison of the gene cluster found in GUM007\_hs to that for aeruginoside 126A (compound 23), with homologous *aer* and *dys* genes highlighted, as well as domain structures delineated. (b) The molecular network includes standards of dysinosins B (compound 26) and C (compound 27) and two new desoxydysinosins (compounds 24 and 25) from GUM007. A small amount of compound 27 was also observed in the sample of GUM007. Domain abbreviations: FkbH, FkbH-like domain; T, thiolation domain; C, condensation domain; A, adenylation domain; E, epimerase domain; KR, ketoreductase domain.

isolated the highly brominated species by preparative HPLC and characterized their structures by comparing tandem mass spectrometry and comprehensive NMR spectroscopy (Text S1) to previously reported PBDEs from *L. herbacea* specimens (8, 52). Notably, one of the aromatic rings remains unchanged as a dibromophenol moiety, while the second phenol ring varies extensively in its bromination patterns in compounds 13 to 22. We suspect that *hs\_bmp18* functions as the additional brominase of the dihydroxy-PBDE 15 to give the pentabrominated 16 and 18 and the hexabrominated 21 and their respective *O*-methylated products. Compounds 13, 14, and 20 are, however, more difficult to rationalize and suggest novel debromination and/or bromine isomerization biochemistry.

**Genome mining leads to discovery of novel dysinosins.** Within the genome of GUM007\_hs, but not GUM202\_hs, we identified an NRPS cluster homologous to the aeruginoside BGC from the cyanobacterium *Planktothrix agardhii* CYA126/8 (50) (Fig. 5a). The GUM007\_hs NRPS cluster contains genes homologous to *aerB*, an NRPS gene with an adenylation (A) domain selective for leucine; *aerC* to -G, genes responsible for the assembly and attachment of the unusual Choi amino acid moiety; and *aerI*, a putative glycosyltransferase. The only notable gene missing is *aerA*, the NRPS loading domain that installs the phenylpyruvate unit in aeruginoside (23). We thus suspected

that the GUM007\_hs *aer*-like BGC (named *dys*) might encode the synthesis of the structurally related dysinosins that primarily differ in their N-terminal unit, a sulfated glyceric acid. The *aerB* homologue in GUM007\_hs, *dysB*, encodes a substantially larger octadomain protein with an unusual sulfotransferase domain consistent with dysinosin's diagnostic sulfated glyceric acid. Additional adenomethyltransferase and FkbH domains in *DysB* are implicated in constructing the sulfated glyceric acid moiety. *DysB* further contains two peptidyl carrier protein (PCP) domains, one condensation (C) domain, an A domain with no consensus specificity predicted by bioinformatic tools, and an epimerization (E) domain. Taken all together, *DysB* is a large multimodular NRPS-type complex that appears to be responsible for the activation of a sulfated glyceric acid and addition of an unspecified  $\beta$ -amino acid.

The logical bioinformatic basis for a putative dysinosin cluster led us to examine GUM007 extracts for dysinosin-like molecules (18, 19). Dysinosins are potent inhibitors of the blood coagulation cascade factor VIIa and the serine protease thrombin (18). We discovered masses representing two new desoxydysinosins (compounds 24 and 25) related in structure to dysinosins B (compound 26) and C (compound 27) containing one hydroxyl group on the Choi moiety instead of two (Fig. 5b). The new dysinosins were found to have masses of 603.2806 (compound 24, 0.17-ppm error from theoretical  $m/z$  603.2807) and 765.3327 (compound 25, 1.04-ppm error from theoretical  $m/z$  765.3335). Interrogation of the MS/MS data revealed that in each of the dysinosin standards 26 and 27, a neutral mass loss is observed, which is consistent with the loss of sulfate from the terminal glyceric acid residue. We observed the same neutral mass loss in the new GUM007 dysinosins 24 (Fig. S6) and 25 (Fig. S7). GNPS molecular networking (53) produced a network of parent masses and corresponding neutral mass loss for both standards and the two new dysinosins (Fig. 5b). The combined MS/MS data, molecular networking, and genomic information support the discovery of two new dysinosin molecules using genome-mining-guided isolation. Understanding the biosynthetic basis of these bioactive molecules may provide a platform for rational molecule design for more effective and selective inhibitors in the blood coagulation cascade (18, 19).

**Conclusions.** Marine sponges continue to be one of the most promising sources of marine natural products, but challenges remain in obtaining meaningful amounts of a molecule and identifying the true producer of a given natural product. Physically vulnerable sponges are thought to rely on their microbial communities for specialized chemical defense, making the sponge microbiome a promising, untapped source of new natural products. Until recently, the full biosynthetic potential of uncultured symbionts remained concealed in their inaccessible genomes, but advances in sequencing read length and metagenomic assembly are beginning to reveal symbionts as important subjects for genome mining efforts.

The high-quality draft genomes generated for two *Hormoscilla* populations afforded a complete look at the biosynthetic capacity of these uncultivated symbionts and represent the strongest evidence yet for cyanobacterial production of multiple classes of compounds isolated from sponges, including the PBDEs and dysinosins. The variety of secondary metabolite classes encoded in the genome of just one symbiont presents a concise strategy for the sponge host to gain chemical diversity while maintaining a relationship with one dominant bacterial strain, rather than a diversified community of secondary metabolite-producing bacteria. This genomic information identified a new type of brominating enzyme that appears to directly brominate hydroxy-PBDEs and also led to the discovery of two new dysinosins and the genetic basis to begin understanding the biosynthesis of this bioactive family of compounds.

The high-quality draft genomes also allowed for reliable comparisons between *Hormoscilla* population genomes and well-characterized free-living cyanobacteria, exposing multiple avenues for exploring these symbionts' recalcitrance to cultivation in the lab. We observed incomplete pathways of essential primary metabolites, including histidine, thiamine, and biotin. Although they are as yet unable to be cultured in the

lab, these symbionts defy the strategy taken by many obligate symbionts and retain a large genome. Their expanded chemical repertoire may discourage large-scale genome streamlining. Obtaining high-quality genomes from metagenomes is the next frontier in microbiology and will begin to populate sequence databases with uncultivated strains, ultimately giving us a better look into their symbiotic lives and their yet-unseen potential for bioactive metabolite production.

## MATERIALS AND METHODS

**Hybrid sequencing and assembly of enriched *Hormoscilla* populations.** GUM007 and GUM202 were collected in July 2015 by snorkel in Guam in Pago Bay (8) and in December 2016 by scuba diving in 20 to 40 feet of water at Anae Island, respectively. Both sponges were processed to obtain enriched cyanobacterial fractions, according to a trichome enrichment protocol described previously (8). The enriched cyanobacterial fractions were subjected to genomic DNA isolation using a previously published protocol (54) to obtain high-molecular-weight DNA. Preparation of all libraries and sequencing were performed by the UC San Diego Institute for Genomic Medicine. PacBio libraries were constructed from the enriched, high-molecular-weight DNA for both GUM007 and GUM202 (see Fig. S1 in the supplemental material). PacBio sequencing libraries were generated using SMRTbell template preparation reagent kits (Pacific Biosciences), and libraries of >6 kb were selected using a PippinHT (Sage Science). Libraries were sequenced on a PacBio RS II sequencer (UCSD IGM Genomics Center, La Jolla, CA) via 4-h movies using the DNA/polymerase binding kit version P6 V2 with C<sub>4</sub> sequencing chemistry. Unenriched metagenomic DNA was extracted from GUM202 whole-sponge tissue frozen in RNAlater as previously described (8). Illumina TruSeq libraries were constructed from bulk metagenomic sponge DNA for GUM202 and enriched trichome DNA for GUM007 and sequenced on an Illumina HiSeq 2500 using a 2-by 150-bp paired-end sequencing strategy.

The GUM007\_hs and GUM202\_hs genomes were assembled using two slightly different methods. The GUM007 genome was assembled as follows: paired-end Illumina reads were quality filtered and trimmed using Trimmomatic version 0.35 (55), with the parameters LEADING:3, TRAILING:3, HEADCROP:9, SLIDINGWINDOW:4:15, MINLEN:100, and then assembled with IDBA-UD version 1.1.1 set to default parameters (56). Preliminary contigs were grouped into bins based on percent GC, nucleotide composition, assembly depth of coverage, and taxonomic assignment by DarkHorse version 1.5 (57), as previously described (58). Reads were mapped to contigs classified as cyanobacteria using the end-to-end mode of Bowtie2, version 2.2.18 (59). Coverage depth was calculated using the idxstats module of SAMtools version 0.1.191 (60). Read subsets from scaffold bins identified as potentially belonging to *Cyanobacteria* were reassembled using Celera Assembler version 8.3 (61), configured with merSize = 18, utgGenomeSize = 6 Mb, and utgErrorRate = 0.02. PacBio reads were recruited to scaffolds from this assembly based on blastn E values of 1e<sup>-7</sup> or better, filtered for a minimum size of 500 nt or greater, and then used to scaffold the Illumina assembled genome into longer sequences using SSPACE-LongRead (62).

The GUM202\_hs genome was obtained by first assembling paired 150-bp Illumina reads from the GUM202 *L. herbacea* metagenome with IDBA-UD, using read trimming and assembly parameters as described above. All scaffolds were taxonomically assigned using the same strategy as described in reference 58 but employing DarkHorse version 2, which used a database customized to include the GUM007 genome. *Hormoscilla* scaffolds were binned based on shared taxonomic assignment, sequence composition, and read coverage information in Anvi'o, version 2.2.2 (63). *Hormoscilla*-associated reads were identified by mapping the Illumina data set to this bin using the local mode of Bowtie2, version 2.2.7. The PacBio reads were also phylogenetically associated with cyanobacteria by BLASTing against a custom-made database of protein-coding sequences from the GUM007\_hs genome previously assembled and a database made of protein-coding sequences for all *Oscillatoria* genomes in the Joint Genome Institute (JGI) database (64). Any PacBio reads that contained a hit to one of these databases was retained. The raw, binned Illumina reads and the raw positive-hit PacBio reads were then used for assembly with hybridSPAdes in SPAdes version 3.10.1 using standard parameters (65). The Illumina reads and the GUM202\_hs hybrid assembly were then used as input for GapFiller to obtain the final assembly (66).

Quality and completeness of genome assemblies were determined using CheckM version 1.0.11 using the lineage workflow (31). Average nucleotide identity (ANI) and average amino acid identity (AAI) were calculated by submitting both genome assemblies and the shared gene clusters to the ANI/AAI calculator at <http://enve-omics.ce.gatech.edu/>. The genome synteny plot was created using the MUMmer version 4.0.0 suite, including NUCmer for alignment, delta-filter for filtering, and mummerplot for visualization using the -fat and -large options (67). Gene cluster synteny images were created using MultiGeneBlast (68).

**Phylogenetic analysis.** A multilocus sequence analysis (MLSA) was used to build a phylogenetic tree using 25 conserved housekeeping genes (Table S2) from a set of 31 genes previously defined for bacterial MLSA (69). All 442 available cyanobacteria genomes in JGI as of 22 January 2018 were searched for the 25 housekeeping genes, of which 305 genomes contained each gene in one copy. After eliminating duplicate genomes and limiting redundant species to five representatives, a final set of 197 cyanobacterial genomes were used in the MLSA (Text S1). *Chloroflexus aurantiacus* J-10-fl was used as an outgroup. Each of the 25 gene sets was individually aligned using MAFFT version 7.310 (70) with high-accuracy local iterative mode using 100 iterations. Next, each alignment was trimmed using trimAl version 1.2rev59 (71) and the "automated1" option optimized for maximum-likelihood tree construction. The resulting

trimmed, aligned files were concatenated using a custom Python script, and the resulting supermatrix was processed with IQ-TREE version 1.6.1 (72) with 1,000 ultrabootstrap replicates using UF:Boot2 (73) and ModelFinder (74) for each gene partition. The tree was visualized using interactive tree of life (iTOL) version 3 (75).

**Primary metabolism and secondary metabolism analyses.** The primary metabolism analysis consisted of an initial BLAST of *Synechococcus elongatus* PCC 7942 protein-coding regions against both GUM\_hs genomes with an E value cutoff of  $e^{-20}$ . Missing genes were converted into their corresponding KEGG Ontology (KO) and pfam numbers for analysis in JGI IMG/MER. KOs and pfams that were absent in both *Hormoscilla* genomes were cross-referenced against gene essentiality in *S. elongatus*. The intersection of the two data sets was a set of seven genes missing in the GUM\_hs genomes and essential in *S. elongatus* PCC 7942 (Table S1). In order to determine if the pathways containing missing genes were actually incomplete, ec2kegg (76) was used to generate metabolic maps for pathway comparison in *S. elongatus* PCC 7942 and the GUM\_hs genomes (Fig. S3 and S4).

For secondary metabolism analysis, the two genomes were submitted to antiSMASH 4.1.0 with standard options (77). Biosynthetic gene cluster networking was done using a locally installed version of the BiG-SCAPE software with the local option enabled (46). The resulting pairwise scores were filtered for those above 0.40 and were visualized as a network using Gephi version 0.9.1 (78).

**Chemical structure elucidations.** Lyophilized GUM202 sponge tissue (1.5 g) was pulverized and extracted with  $3 \times 20$ -ml methanol (MeOH) for 30 to 60 min on a benchtop nutator. The combined extracts were dried *in vacuo*, resuspended in dichloromethane ( $\text{CH}_2\text{Cl}_2$ ), and partitioned against water. The resulting  $\text{CH}_2\text{Cl}_2$  layer was dried using magnesium sulfate, filtered, and dried *in vacuo*. Preparative HPLC solvents were HPLC-grade water with 0.1% trifluoroacetic acid (TFA) and HPLC-grade acetonitrile (MeCN) with 0.1% TFA. An Agilent 218 purification system (ChemStation software; Agilent) equipped with a ProStar 410 automatic injector, Agilent ProStar UV-Vis dual-wavelength detector, 440-LC fraction collector, and an Agilent Pursuit XRs 5  $\text{C}_{18}$  100- by 21.2-mm preparative HPLC column was used (0 to 5 min 5% MeCN isocratic, 5 to 10 min 10 to 25% MeCN, 10 to 35 min 65 to 75% MeCN, 35 to 40 min 75 to 100% MeCN, and 40 to 45 min 100% MeCN isocratic). Ten fractions were collected and analyzed using LC-MS/MS performed on an Agilent 1260 LC system with diode array detector and a Phenomenex Kinetex 5  $\mu\text{C}_{18}$ (2) 100-A, 150- by 4.6-mm column in negative mode. LC-MS solvents were LC-MS-grade water with 0.1% formic acid and LC-MS-grade acetonitrile with 0.1% formic acid. Mass spectra were analyzed with Agilent MassHunter qualitative analysis version B.05.00. Purified PBDEs were validated by 1D and 2D NMR on a JEOL 500-MHz NMR spectrometer in deuterated methanol. Chemical shift trends were matched to previously reported PBDEs (52).

For the desoxydysinosins, GUM007 whole-sponge tissue was pulverized and extracted with  $3 \times 20$ -ml MeOH for 30 to 60 min on a benchtop nutator. The combined extracts were dried *in vacuo*. Approximately 100 mg MeOH extract and 40 ml RNAlater solution used to store the sponges were loaded onto 1-g  $\text{C}_{18}$  solid-phase extraction columns (Canadian Life Science) and fractionated from 5% to 100% MeCN. Three standards, dysinosin A (compound 10), dysinosin B (compound 27), and dysinosin C (compound 28), kindly provided by Ron Quinn from Griffith University, were prepared as 0.1-mg/ml solutions in MeOH and analyzed with the GUM007 MeCN fractions using the Agilent 6530 Accurate-Mass Q-TOF MS mentioned above with a Phenomenex Kinetex 5  $\mu\text{C}_{18}$ (2) 100-A, 150- by 4.6-mm column (0 to 3 min 5% MeCN isocratic, 3 to 23 min 5 to 100% MeCN, 23 to 26 min 100% MeCN isocratic at 0.7 ml/min). The dysinosin standards revealed a characteristic loss of sulfate, which led to the discovery of the desoxydysinosins, present in the 15% and 20% MeCN fractions. Molecular networking, using the GNPS platform (53), was used to visualize the dysinosin molecules.

**Data availability.** The GUM007 Whole-Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession numbers RFFB00000000, BioSample SAMN10266268, and BioProject PRJNA320446. The version described in this paper is RFFB01000000. The GUM202 Whole-Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession numbers RFFC00000000, BioSample SAMN10266269, and BioProject PRJNA320446. The version described in this paper is RFFC01000000.

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mBio.00821-19>.

**TEXT S1**, PDF file, 0.1 MB.

**FIG S1**, PDF file, 0.3 MB.

**FIG S2**, PDF file, 0.3 MB.

**FIG S3**, PDF file, 0.03 MB.

**FIG S4**, PDF file, 0.05 MB.

**FIG S5**, PDF file, 0.1 MB.

**FIG S6**, PDF file, 0.2 MB.

**FIG S7**, PDF file, 0.1 MB.

**TABLE S1**, PDF file, 0.6 MB.

**TABLE S2**, PDF file, 0.4 MB.

## ACKNOWLEDGMENTS

We acknowledge Mohammad Alanjary (University of Tuebingen) for helpful advice with phylogenetic analysis and gene cluster networking, Neha Garg (Georgia Tech) for insights shared about LC-MS characterization of the dysinosins, Jared Broddrick and David Welkie (UCSD) for their assistance in metabolic model comparisons and helpful discussions, Valerie Paul (Smithsonian) and Kate Bauman (UCSD) for their help in collecting sponges in Guam, Ron Quinn (Griffith University) for providing dysinosin standards, and Tristan de Rond (UCSD) for insightful discussions that led to the identification of new dysinosins.

This work was supported by grants from the U.S. National Institutes of Health (R01-ES030316 to B.S.M. and E.E.A., R00-ES026620 to V.A.) and the National Science Foundation (OCE-1837116 to B.S.M. and E.E.A.) and a Sloan Foundation research fellowship (to V.A.).

## REFERENCES

- Newman DJ, Cragg GM. 2016. Natural products as sources of new drugs from 1981 to 2014. *J Nat Prod* 79:629–661. <https://doi.org/10.1021/acs.jnatprod.5b01055>.
- Mehub MF, Perkins MV, Zhang W, Franco C. 2016. New marine natural products from sponges (Porifera) of the order Dictyoceratida (2001 to 2012); a promising source for drug discovery, exploration and future prospects. *Biotechnol Adv* 34:473–491. <https://doi.org/10.1016/j.biotechadv.2015.12.008>.
- Paul VJ, Puglisi MP. 2004. Chemical mediation of interactions among marine organisms. *Nat Prod Rep* 21:189–209. <https://doi.org/10.1039/b302334f>.
- Harvey AL, Edrada-Ebel R, Quinn RJ. 2015. The re-emergence of natural products for drug discovery in the genomics era. *Nat Rev Drug Discov* 14:111–129. <https://doi.org/10.1038/nrd4510>.
- Sharma GM, Vig B. 1972. Studies on the antimicrobial substances of sponges. VI. Structures of two antibacterial substances isolated from the marine sponge *Dysidea herbacea*. *Tetrahedron Lett* 13:1715–1718. [https://doi.org/10.1016/S0040-4039\(01\)84729-8](https://doi.org/10.1016/S0040-4039(01)84729-8).
- Carté B, Faulkner DJ. 1981. Polybrominated diphenyl ethers from *Dysidea herbacea*, *Dysidea chlorea* and *Phyllospongia foliascens*. *Tetrahedron* 37:2335–2339. [https://doi.org/10.1016/S0040-4020\(01\)88886-4](https://doi.org/10.1016/S0040-4020(01)88886-4).
- Unson MD, Holland ND, Faulkner DJ. 1994. A brominated secondary metabolite synthesized by the cyanobacterial symbiont of a marine sponge and accumulation of the crystalline metabolite in the sponge tissue. *Mar Biol* 119:9–11. <https://doi.org/10.1007/BF00350100>.
- Agarwal V, Blanton JM, Podell S, Taton A, Schorn MA, Busch J, Lin Z, Schmidt EW, Jensen PR, Paul VJ, Biggs JS, Golden JW, Allen EE, Moore BS. 2017. Metagenomic discovery of polybrominated diphenyl ether biosynthesis by marine sponges. *Nat Chem Biol* 13:537–543. <https://doi.org/10.1038/nchembio.2330>.
- Kazlauskas R, Lidgard RO, Wells RJ, Vetter W. 1977. A novel hexachloro-metabolite from the sponge *Dysidea herbacea*. *Tetrahedron Lett* 18:3183–3186. [https://doi.org/10.1016/S0040-4039\(01\)83192-0](https://doi.org/10.1016/S0040-4039(01)83192-0).
- Kazlauskas R, Murphy PT, Wells RJ. 1978. A diketopiperazine derived from trichloroleucine from the sponge *Dysidea herbacea*. *Tetrahedron Lett* 19:4945–4948. [https://doi.org/10.1016/S0040-4039\(01\)85777-4](https://doi.org/10.1016/S0040-4039(01)85777-4).
- Flatt P, Gautschi J, Thacker R, Musafija-Girt M, Crews P, Gerwick W. 2005. Identification of the cellular site of polychlorinated peptide biosynthesis in the marine sponge *Dysidea* (*Lamellodysidea*) *herbacea* and symbiotic cyanobacterium *Oscillatoria spongeliae* by CARD-FISH analysis. *Mar Biol* 147:761–774. <https://doi.org/10.1007/s00227-005-1614-9>.
- Unson MD, Faulkner DJ. 1993. Cyanobacterial symbiont biosynthesis of chlorinated metabolites from *Dysidea herbacea* (Porifera). *Experientia* 49:349–353. <https://doi.org/10.1007/BF01923420>.
- Flowers AE, Garson MJ, Webb RI, Dumdei EJ, Charan RD. 1998. Cellular origin of chlorinated diketopiperazines in the dictyoceratid sponge *Dysidea herbacea* (Keller). *Cell Tissue Res* 292:597–607. <https://doi.org/10.1007/s004410051089>.
- Kazlauskas R, Murphy PT, Wells RJ. 1978. A new sesquiterpene from the sponge *Dysidea herbacea*. *Tetrahedron Lett* 19:4949–4950. [https://doi.org/10.1016/S0040-4039\(01\)85778-6](https://doi.org/10.1016/S0040-4039(01)85778-6).
- Sakai R, Suzuki K, Shimamoto K, Kamiya H. 2004. Novel betaines from a Micronesian sponge *Dysidea herbacea*. *J Org Chem* 69:1180–1185. <https://doi.org/10.1021/jo0355045>.
- Isaacs S, Berman R, Kashman Y, Gebreyesus T, Yosief T. 1991. New polyhydroxy sterols, dysidamides, and a dideoxyhexose from the sponge *Dysidea herbacea*. *J Nat Prod* 54:83–91. <https://doi.org/10.1021/np50073a004>.
- Bandaranayake WM, Bemis JE, Bourne DJ. 1996. Ultraviolet absorbing pigments from the marine sponge *Dysidea herbacea*: isolation and structure of a new mycosporine. *Comp Biochem Physiol C Pharmacol Toxicol Endocrinol* 115:281–286. [https://doi.org/10.1016/S0742-8413\(96\)00135-1](https://doi.org/10.1016/S0742-8413(96)00135-1).
- Carroll AR, Pierens GK, Fechner G, De Almeida Leone P, Ngo A, Simpson M, Hyde E, Hooper JN, Bostrom SL, Musil D, Quinn RJ. 2002. Dysinosin A: a novel inhibitor of factor VIIa and thrombin from a new genus and species of Australian sponge of the family Dysideidae. *J Am Chem Soc* 124:13340–13341. <https://doi.org/10.1021/ja020814a>.
- Carroll AR, Buchanan MS, Edser A, Hyde E, Simpson M, Quinn RJ. 2004. Dysinosins B-D, inhibitors of factor VIIa and thrombin from the Australian sponge *Lamellodysidea chlorea*. *J Nat Prod* 67:1291–1294. <https://doi.org/10.1021/np049968p>.
- Crawford JM, Clardy J. 2011. Bacterial symbionts and natural products. *Chem Commun* 47:7559–7566. <https://doi.org/10.1039/c1cc11574j>.
- Schmidt EW, Nelson JT, Rasko DA, Sudek S, Eisen JA, Haygood MG, Ravel J. 2005. Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the cyanobacterial symbiont of *Lissoclinum patella*. *Proc Natl Acad Sci U S A* 102:7315–7320. <https://doi.org/10.1073/pnas.0501424102>.
- Wilson MC, Mori T, Ruckert C, Uria AR, Helf MJ, Takada K, Gernert C, Steffens UA, Heycke N, Schmitt S, Rinke C, Helfrich EJ, Brachmann AO, Gurgui C, Wakimoto T, Kracht M, Crusemann M, Hentschel U, Abe I, Matsunaga S, Kalinowski J, Takeyama H, Piel J. 2014. An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* 506:58–62. <https://doi.org/10.1038/nature12959>.
- Berthold RJ, Borowitzka MA, Mackay MA. 1982. The ultrastructure of *Oscillatoria spongeliae*, the blue-green algal endosymbiont of the sponge *Dysidea herbacea*. *Phycologia* 21:327–335. <https://doi.org/10.2216/i0031-8884-21-3-327.1>.
- Thacker RW, Starnes S. 2003. Host specificity of the symbiotic cyanobacterium *Oscillatoria spongeliae* in marine sponges, *Dysidea* spp. *Mar Biol* 142:643–648. <https://doi.org/10.1007/s00227-002-0971-x>.
- Ridley CP, Bergquist PR, Harper MK, Faulkner DJ, Hooper JNA, Haygood MG. 2005. Speciation and biosynthetic variation in four dictyoceratid sponges and their cyanobacterial symbiont, *Oscillatoria spongeliae*. *Chem Biol* 12:397–406. <https://doi.org/10.1016/j.chembiol.2005.02.003>.
- Hinde R, Pironet F, Borowitzka MA. 1994. Isolation of *Oscillatoria spongeliae*, the filamentous cyanobacterial symbiont of the marine sponge *Dysidea herbacea*. *Mar Biol* 119:99–104. <https://doi.org/10.1007/BF00350111>.
- Thacker RW. 2005. Impacts of shading on sponge-cyanobacteria symbioses: a comparison between host-specific and generalist associations. *Integr Comp Biol* 45:369–376. <https://doi.org/10.1093/icb/45.2.369>.

28. Gomez-Escribano JP, Alt S, Bibb MJ. 2016. Next generation sequencing of actinobacteria for the discovery of novel natural products. *Mar Drugs* 14:E78. <https://doi.org/10.3390/md14040078>.
29. 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>.
30. Bowers RM, Kyrpidis NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Etema TJG, Tighe S, Konstantinidis KT, Liu W-T, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Kyrpidis NC, Schriml L, Garrity GM, Hugenholtz P, Sutton G, Yilmaz P, Meyer F, Glöckner FO, Gilbert JA, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, Schriml L, Banfield JF, Hugenholtz P, Woyke T. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 35:725–731. <https://doi.org/10.1038/nbt.3893>.
31. 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>.
32. Konstantinidis KT, Rosselló-Móra R, Amann R. 2017. Uncultivated microbes in need of their own taxonomy. *ISME J* 11:2399–2406. <https://doi.org/10.1038/ismej.2017.113>.
33. Penn K, Jenkins C, Nett M, Udway DW, Gontang EA, McGlinchey RP, Foster B, Lapidus A, Podell S, Allen EE, Moore BS, Jensen PR. 2009. Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria. *ISME J* 3:1193–1203. <https://doi.org/10.1038/ismej.2009.58>.
34. Meyer JL, Paul VJ, Raymundo LJ, Teplitski M. 2017. Comparative metagenomics of the polymicrobial black band disease of corals. *Front Microbiol* 8:618. <https://doi.org/10.3389/fmicb.2017.00618>.
35. Gao ZM, Wang Y, Tian RM, Wong YH, Batang ZB, Al-Suwailem AM, Bajic VB, Qian PY. 2014. Symbiotic adaptation drives genome streamlining of the cyanobacterial sponge symbiont “Candidatus Synechococcus spongiarum.” *mBio* 5:e00079-14. <https://doi.org/10.1128/mBio.00079-14>.
36. 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>.
37. Donia MS, Fricke WF, Partensky F, Cox J, Elshahawi SI, White JR, Phillippy AM, Schatz MC, Piel J, Haygood MG, Ravel J, Schmidt EW. 2011. Complex microbiome underlying secondary and primary metabolism in the tunicate-Prochloron symbiosis. *Proc Natl Acad Sci U S A* 108:E1423–E1432. <https://doi.org/10.1073/pnas.1111712108>.
38. Lackner G, Peters EE, Helfrich EJ, Piel J. 2017. Insights into the lifestyle of uncultured bacterial natural product factories associated with marine sponges. *Proc Natl Acad Sci U S A* 114:E347–E356. <https://doi.org/10.1073/pnas.1616234114>.
39. Broddrick JT, Rubin BE, Welkie DG, Du N, Mih N, Diamond S, Lee JJ, Golden SS, Palsson BO. 2016. Unique attributes of cyanobacterial metabolism revealed by improved genome-scale metabolic modeling and essential gene analysis. *Proc Natl Acad Sci U S A* 113:E8344–E8353. <https://doi.org/10.1073/pnas.1613446113>.
40. Rubin BE, Wetmore KM, Price MN, Diamond S, Shultzaberger RK, Lowe LC, Curtin G, Arkin AP, Deuschbauer A, Golden SS. 2015. The essential gene set of a photosynthetic organism. *Proc Natl Acad Sci U S A* 112:E6634–E6643. <https://doi.org/10.1073/pnas.1519220112>.
41. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 44:D457–D462. <https://doi.org/10.1093/nar/gkv1070>.
42. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heeger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M. 2014. Pfam: the protein families database. *Nucleic Acids Res* 42:D222–D230. <https://doi.org/10.1093/nar/gkt1223>.
43. Leao T, Castela G, Korobeynikov A, Monroe EA, Podell S, Glukhov E, Allen EE, Gerwick WH, Gerwick L. 2017. Comparative genomics uncovers the prolific and distinctive metabolic potential of the cyanobacterial genus *Moorea*. *Proc Natl Acad Sci U S A* 114:3198–3203. <https://doi.org/10.1073/pnas.1618556114>.
44. Cummings SL, Barbe D, Leao TF, Korobeynikov A, Engene N, Glukhov E, Gerwick WH, Gerwick L. 2016. A novel uncultured heterotrophic bacterial associate of the cyanobacterium *Moorea* produces JHB. *BMC Microbiol* 16:198. <https://doi.org/10.1186/s12866-016-0817-1>.
45. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de los Santos E, Kim HU, Nave M, Dickschat JS, Mitchell DA, Shelest E, Breitling R, Takano E, Lee SY, Weber T, Medema MH. 2017. antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res* 45:W36–W41. <https://doi.org/10.1093/nar/gkx319>.
46. Navarro-Muñoz J, Selem-Mojica N, Mallowney M, Kautsar S, Tryon J, Parkinson E, De Los Santos E, Yeong M, Cruz-Morales P, Abubucker S, Roeters A, Lokhorst W, Fernandez-Guerra A, Dias Cappelini LT, Thomson R, Metcalf W, Kelleher N, Barona-Gomez F, Medema MH. 2018. A computational framework for systematic exploration of biosynthetic diversity from large-scale genomic data. *bioRxiv* <https://doi.org/10.1101/445270>.
47. Balskus EP, Walsh CT. 2010. The genetic and molecular basis for sun-screen biosynthesis in cyanobacteria. *Science* 329:1653–1656. <https://doi.org/10.1126/science.1193637>.
48. Klisch M, Häder DP. 2008. Mycosporine-like amino acids and marine toxins—the common and the different. *Mar Drugs* 6:147–163. <https://doi.org/10.3390/md20080008>.
49. Calteau A, Fewer DP, Latifi A, Coursin T, Laurent T, Jokela J, Kerfeld CA, Sivonen K, Piel J, Gugger M. 2014. Phylum-wide comparative genomics unravel the diversity of secondary metabolism in cyanobacteria. *BMC Genomics* 15:977. <https://doi.org/10.1186/1471-2164-15-977>.
50. Ishida K, Christiansen G, Yoshida WY, Kurmayer R, Welker M, Valls N, Bonjoch J, Hertweck C, Borner T, Hemscheidt T, Dittmann E. 2007. Biosynthesis and structure of aeruginoside 126A and 126B, cyanobacterial peptide glycosides bearing a 2-carboxy-6-hydroxyoctahydroindole moiety. *Chem Biol* 14:565–576. <https://doi.org/10.1016/j.chembiol.2007.04.006>.
51. Agarwal V, El Gamal AA, Yamanaka K, Poth D, Kersten RD, Schorn M, Allen EE, Moore BS. 2014. Biosynthesis of polybrominated aromatic organic compounds by marine bacteria. *Nat Chem Biol* 10:640–647. <https://doi.org/10.1038/nchembio.1564>.
52. Calcul L, Chow R, Oliver AG, Tenney K, White KN, Wood AW, Fiorilla C, Crews P. 2009. NMR strategy for unraveling structures of bioactive sponge-derived oxy-polyhalogenated diphenyl ethers. *J Nat Prod* 72:443–449. <https://doi.org/10.1021/np800737z>.
53. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu W-T, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu C-C, Floros DJ, Gavilan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw C-C, Yang Y-L, Humpf H-U, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, Larson CB, Boya P CA, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque DP, Pociute E, O’Neill EC, Briand E, Helfrich EJM, Granatosky EA, Glukhov E, Ryffel F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodríguez AMC, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard P-M, Phapale P, Nothias L-F, Alexandrov T, Litaudon M, Wolfender J-L, Kyle JE, Metz TO, Peryea T, Nguyen D-T, VanLeer D, Shinn P, Jadhav A, Müller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson BO, Pogliano K, Lington RG, Gutiérrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein PC, Bandeira N. 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* 34:828–837. <https://doi.org/10.1038/nbt.3597>.
54. Schmidt EW, Donia MS. 2009. Cyanobactin ribosomally synthesized peptides—a case of deep metagenome mining. *Methods Enzymol* 458:575–596. [https://doi.org/10.1016/S0076-6879\(09\)04823-X](https://doi.org/10.1016/S0076-6879(09)04823-X).
55. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
56. 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. <https://doi.org/10.1093/bioinformatics/bts174>.
57. Podell S, Gaasterland T. 2007. DarkHorse: a method for genome-wide prediction of horizontal gene transfer. *Genome Biol* 8:R16. <https://doi.org/10.1186/gb-2007-8-2-r16>.
  58. Podell S, Ugalde JA, Narasingarao P, Banfield JF, Heidelberg KB, Allen EE. 2013. Assembly-driven community genomics of a hypersaline microbial ecosystem. *PLoS One* 8:e61692. <https://doi.org/10.1371/journal.pone.0061692>.
  59. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
  60. Bascom-Slack CA, Ma C, Moore E, Babbs B, Fenn K, Greene JS, Hann BD, Keehner J, Kelley-Swift EG, Kembaiyan V, Lee SJ, Li P, Light DY, Lin EH, Schorn MA, Vekhter D, Boulanger LA, Hess WM, Vargas PN, Strobel GA, Strobel SA. 2009. Multiple, novel biologically active endophytic actinomycetes isolated from upper Amazonian rainforests. *Microb Ecol* 58:374–383. <https://doi.org/10.1007/s00248-009-9494-z>.
  61. Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, Kravitz SA, Mobarry CM, Reinert KH, Remington KA, Anson EL, Bolanos RA, Chou HH, Jordan CM, Halpern AL, Lonardi S, Beasley EM, Brandon RC, Chen L, Dunn PJ, Lai Z, Liang Y, Nusskern DR, Zhan M, Zhang Q, Zheng X, Rubin GM, Adams MD, Venter JC. 2000. A whole-genome assembly of *Drosophila*. *Science* 287:2196–2204. <https://doi.org/10.1126/science.287.5461.2196>.
  62. Boetzer M, Pirovano W. 2014. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics* 15:211. <https://doi.org/10.1186/1471-2105-15-211>.
  63. Eren AM, Esen OC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 3:e1319. <https://doi.org/10.7717/peerj.1319>.
  64. Markowitz VM, Chen IM, 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>.
  65. Antipov D, Korobeynikov A, McLean JS, Pevzner PA. 2016. hybridSPAdes: an algorithm for hybrid assembly of short and long reads. *Bioinformatics* 32:1009–1015. <https://doi.org/10.1093/bioinformatics/btv688>.
  66. Nadalin F, Vezzi F, Policriti A. 2012. GapFiller: a de novo assembly approach to fill the gap within paired reads. *BMC Bioinformatics* 13(Suppl 14):S8. <https://doi.org/10.1186/1471-2105-13-S14-S8>.
  67. Marcais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput Biol* 14:e1005944. <https://doi.org/10.1371/journal.pcbi.1005944>.
  68. Medema MH, Takano E, Breitling R. 2013. Detecting sequence homology at the gene cluster level with MultiGeneBlast. *Mol Biol Evol* 30:1218–1223. <https://doi.org/10.1093/molbev/mst025>.
  69. Wu M, Eisen JA. 2008. A simple, fast, and accurate method of phylogenomic inference. *Genome Biol* 9:R151. <https://doi.org/10.1186/gb-2008-9-10-r151>.
  70. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>.
  71. 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>.
  72. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>.
  73. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* 35:518–522. <https://doi.org/10.1093/molbev/msx281>.
  74. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jeremiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589. <https://doi.org/10.1038/nmeth.4285>.
  75. Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44:W242–W245. <https://doi.org/10.1093/nar/gkw290>.
  76. Porollo A. 2014. EC2KEGG: a command line tool for comparison of metabolic pathways. *Source Code Biol Med* 9:19. <https://doi.org/10.1186/1751-0473-9-19>.
  77. Weber T, Blin K, Duddela S, Krug D, Kim HU, Brucoleri R, Lee SY, Fischbach MA, Müller R, Wohlleben W, Breitling R, Takano E, Medema MH. 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res* 43:W237–W243. <https://doi.org/10.1093/nar/gkv437>.
  78. Bastian M, Heymann S, Jacomy M. 2009. Gephi: an open source software for exploring and manipulating networks. *Int AAAI Conf Weblogs Social Media*.
  79. Moya A, Pereto J, Gil R, Latorre A. 2008. Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat Rev Genet* 9:218–229. <https://doi.org/10.1038/nrg2319>.