AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems®

RESEARCH ARTICLE

Comparative Metagenomic Analysis of Biosynthetic Diversity across Sponge Microbiomes Highlights Metabolic Novelty, Conservation, and Diversification

Catarina Loureiro,[a] Anastasia Galani,[a] Asimenia Gavriilidou,[a] Maryam Chaib de Mares,[b] John van der Oost,[a] Marnix H. Medema,[c] Detmer Sipkema[a]

[a]Laboratory of Microbiology, Wageningen University, Wageningen, the Netherlands
[b]Groningen Institute for Evolutionary Life Sciences (GELIFES), University of Groningen, Groningen, the Netherlands
[c]Bioinformatics Group, Wageningen University, Wageningen, the Netherlands

ABSTRACT Marine sponges and their microbial symbiotic communities are rich sources of diverse natural products (NPs) that often display biological activity, yet little is known about the global distribution of NPs and the symbionts that produce them. Since the majority of sponge symbionts remain uncultured, it is a challenge to characterize their NP biosynthetic pathways, assess their prevalence within the holobiont, and measure the diversity of NP biosynthetic gene clusters (BGCs) across sponge taxa and environments. Here, we explore the microbial biosynthetic landscapes of three high-microbial-abundance (HMA) sponges from the Atlantic Ocean and the Mediterranean Sea. This data set reveals striking novelty, with <1% of the recovered gene cluster families (GCFs) showing similarity to any characterized BGC. When zooming in on the microbial communities of each sponge, we observed higher variability of specialized metabolic and taxonomic profiles between sponge species than within species. Nonetheless, we identified conservation of GCFs, with 20% of sponge GCFs being shared between at least two sponge species and a GCF core comprised of 6% of GCFs shared across all species. Within this functional core, we identified a set of widespread and diverse GCFs encoding nonribosomal peptide synthetases that are potentially involved in the production of diversified ether lipids, as well as GCFs putatively encoding the production of highly modified proteusins. The present work contributes to the small, yet growing body of data characterizing NP landscapes of marine sponge symbionts and to the cryptic biosynthetic potential contained in this environmental niche.

IMPORTANCE Marine sponges and their microbial symbiotic communities are a rich source of diverse natural products (NPs). However, little is known about the sponge NP global distribution landscape and the symbionts that produce them. Here, we make use of recently developed tools to perform untargeted mining and comparative analysis of sponge microbiome metagenomes of three sponge species in the first study considering replicate metagenomes of multiple sponge species. We present an overview of the biosynthetic diversity across these sponge holobionts, which displays extreme biosynthetic novelty. We report not only the conservation of biosynthetic and taxonomic diversity but also a core of conserved specialized metabolic pathways. Finally, we highlight several novel GCFs with unknown ecological function, and observe particularly high biosynthetic potential in *Acidobacteriota* and *Latescibacteria* symbionts. This study paves the way toward a better understanding of the marine sponge holobionts' biosynthetic potential and the functional and ecological role of sponge microbiomes.

KEYWORDS bioinformatics, marine microbiology, marine sponge, metagenomics, natural products

Editor Laura M. Sanchez, University of California—Santa Cruz

Address correspondence to Detmer Sipkema, detmer.sipkema@wur.nl, or Marnix H. Medema, marnix.medema@wur.nl.

The authors declare a conflict of interest. M.H.M. is a co-founder of Design Pharmaceuticals and a member of the scientific advisory board of Hexagon Bio.

Received 14 April 2022
Accepted 21 June 2022
Published 18 July 2022

Marine sponges (*Porifera*) are benthic heterotrophic filter feeders that harbor diverse and abundant communities of microbial symbionts in their tissues (1, 2). These communities have aided marine sponges in their expansion across diverse ecological niches (2–4), and are often dominated by *Proteobacteria*, *Acidobacteria*, and *Chloroflexi*, as well as the sponge-specific *Poribacteria* (5–7). The complex unit of a sponge and its microbial consortium is referred to as a "holobiont" (8) and can be divided into two categories based on the abundance and diversity of microbes in the sponge tissue, with high-microbial-abundance (HMA) sponges harboring richer and more diverse communities (7, 9–11) at 10^8 to 10^{10} microbial cells g^{-1} sponge wet weight (5, 7) and low-microbial-abundance (LMA) sponges hosting on average 10^5 to 10^6 microbes g^{-1} sponge wet weight (9). Although there are exceptions to this rule (12), there is commonly an enrichment of *Poribacteria*, *Chloroflexi*, and *Acidobacteria* in HMA sponges (11). In addition, there appears to be a functional microbial core that displays gene abundance differences in core metabolic functions (13, 14).

Natural products (NPs) are ubiquitous small molecules that play a key role in symbiosis as mediators of interactions within the holobiont (15, 16). In this generally mutualistic relationship, the microbes provide their host with primary nutrients and chemical compounds that prevent predation, fouling, and infection, while receiving primary metabolic nutrition and a hospitable habitat (17–22). While symbiosis often leads to the reduction and specialization of a symbiont's genome, the sponge holobiont specialized metabolism is maintained through positive selective pressure (23, 24). This specialization can also lead to the generation of "super producers" with high numbers of biosynthetic gene clusters, such as the genus "*Candidatus* Enthotheonella," that was first discovered in sponges (15, 25). Sponge symbionts are a particularly prolific source of diverse NPs that often display biological activity (3, 26, 27). However, the large majority of sponge symbionts remain uncultured, which has hindered characterization of NP-mediated host-symbiont interactions (28, 29) and, consequently, access to this untapped reservoir of a broad spectrum of bacterial specialized metabolites (30, 31).

The enzymes that catalyze the production of specific specialized metabolites are generally encoded in biosynthetic gene clusters (BGCs) (32). BGCs discovered in marine symbiotic systems display a high degree of diversity, with noncanonical cluster architectures that allow for the biosynthesis of highly diverse specialized metabolites (25, 30, 33, 34). Still, there are a relatively few studies examining the sponge holobiont with a focus on specialized metabolism (30, 35, 36). Metagenomics aids in recovering environmental genetic material from uncultured microbes, and reconstruction of metagenome-assembled genomes (MAGs) facilitates studying these gene clusters within their genomic and taxonomic context (37–39). Recent development of tools such as antiSMASH (40) and BiG-SCAPE (41) (the most widely used tools for prediction and comparison of BGCs), as well as MIBiG (32; a curated database for BGCs with experimentally determined products), allow for efficient leveraging of metagenomic sequencing data sets for discovery of BGCs. This approach has revealed the existence of uncharacterized microbes with diverse specialized metabolism repertoires (42, 43) and facilitated linking putative specialized metabolites to their bacterial producers (31, 33, 44).

Untargeted mining of sponge microbiome metagenomes allows for a more complete view of the holobiont's biosynthetic diversity and conservation across sponge holobionts (15). While NPs have mostly been described as highly niche-specialized metabolites, there are emerging data pointing to the conservation of certain BGCs in a symbiotic context (15, 45). A broader view on the extent to which BGCs are shared between the symbiotic communities of different sponge species, however, is still missing. Here, we make use of these culture-independent methods to explore the combined taxonomic and biosynthetic landscape of select marine sponge bacterial symbiont communities. In this way, we establish a detailed overview of specialized metabolic diversity in these sponge holobiont, which comprises a diverse array of mostly uncharacterized gene cluster families (GCFs). The ubiquity of a part of this array supports the hypothesis of an ecologically important set of specialized metabolic pathways conserved among the bacterial symbionts of HMA

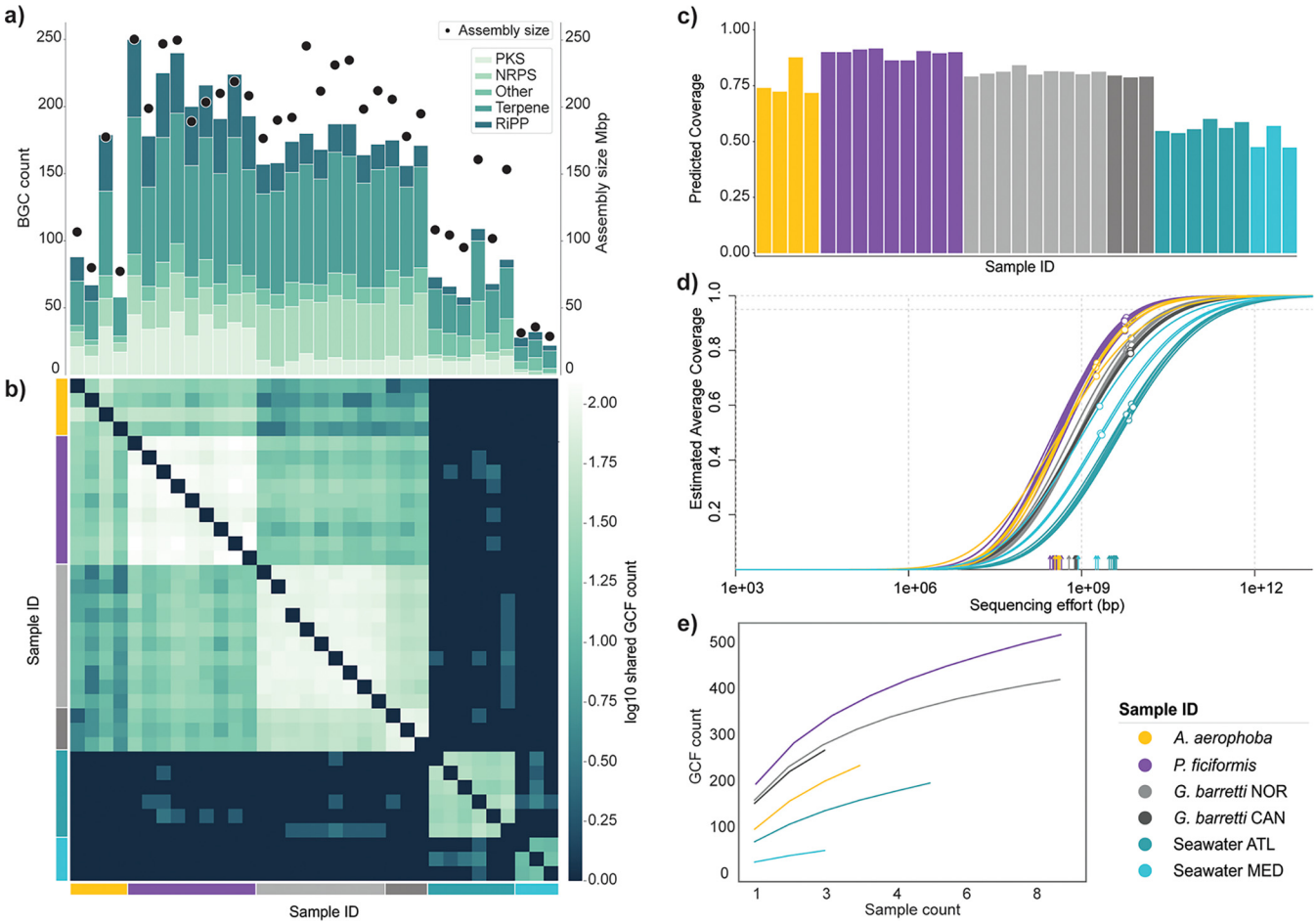

FIG 1 BGCs and GCFs in marine sponges. (a) BGC counts colored per major class (right *y*-axis, stacked bars) and assembly size (left *y*-axis, black dots) per sample. Assembly size includes only contigs >4,000 bp in length, used as input for antiSMASH. (b) Log_{10}-normalized pairwise heatmap of shared GCF counts between samples. (c) Nonpareil sequence coverage estimates. (d) Nonpareil curves estimating the relationship between estimated coverage (*y*-axis; white dots indicate sample estimated coverage) and sequencing effort (*x*-axis; vertical arrows indicate sample sequencing effort). (e) GCF rarefaction curves.

sponge species. This includes a novel set of nonribosomal peptide synthetase (NRPS)-like GCFs associated with the production of molecules related to vinyl ether lipid phosphatidylethanolamine (VEPE) (46), as well as multiple GCFs spanning several ribosomally synthesized and posttranslationally modified peptide (RiPP) classes that appear unique to sponge microbiota. In addition, we identified the putative bacterial hosts of these BGCs, which constitutes an important next step toward accessing the biosynthetic potential that is still largely untapped in sponge microbiomes.

RESULTS AND DISCUSSION

Sponges harbor a striking number of novel biosynthetic gene clusters. In total, 5,082 BGCs were detected in the sponge and seawater samples, which were grouped into 1,186 GCFs plus 394 singletons. Of all the recovered GCFs, only four included experimentally characterized reference BGCs from MIBiG (47), a striking display of the potential novelty contained within marine sponges. We can observe differences between the bacterial communities of all sponge species with regard to their BGC counts, as well as sequencing depth and assembly size (Fig. 1a). The extended period between samplings and the diverse nature of the sequencing performed for each sample set warranted looking into a possible impact of the type of sequencing performed on sequence coverage and genomic content recovery. Although BGC count patterns and assembly size/sequencing effort do follow the Nonpareil estimated coverage, the latter shows less amplitude in

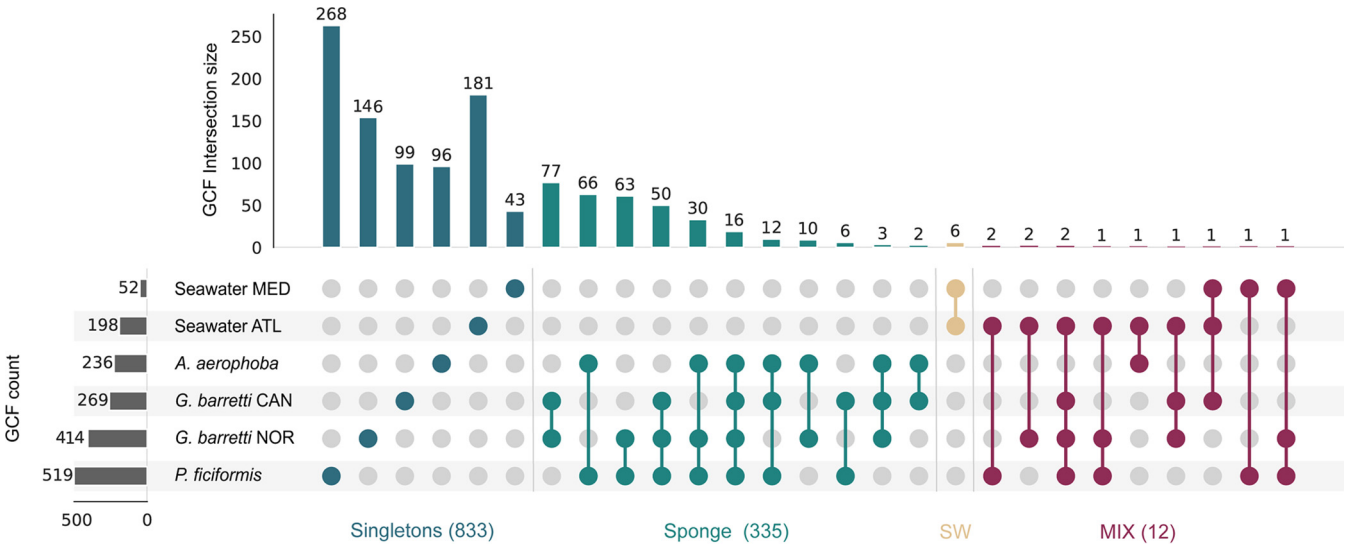

FIG 2 GCF intersections across sample groups. AN UpsetPlot illustrates GCF intersections between sample groups, i.e., samples grouped by sponge species and sampling geographical location. The GCF count refers to the total count for each sample group, and the intersection size refers to the GCF count of each fraction/intersection (grid dots). For each category (singletons, sponge, mix of sponge and seawater) with >1 intersection, the total GCF count is shown in parentheses.

variation (Fig. 1c and d). This indicates that BGC count patterns (Fig. 1a) observed are largely representative of the samples' inherent sequence diversity.

Since BGC boundaries are notoriously difficult to define and BGCs are often fragmented, we have considered the GCF here to be the biosynthetic unit of study in order to minimize inflation originating from fragmented BGCs. Specialized metabolic similarity between samples can then be estimated based on the fraction of shared GCFs (Fig. 1b), with sponge samples showing high similarity even across species and habitat locations, but not to seawater samples (Fig. 1b). GCF discovery rarefaction curves (Fig. 1e) show that the biosynthetic diversity in the HMA sponge holobiont is not fully captured and indicate that a higher sequencing effort would be needed for complete coverage. Nonpareil projection curves (Fig. 1d) estimate the required sequencing effort at 10 to 100 Gbp for 90% sequence diversity recovery for HMA sponge samples. However, with estimated coverage values above 75% for all sponge samples, our data recover a large fraction of this diversity, allowing characterization of the majority of the biosynthetic landscape of these holobionts.

When investigating the distribution of GCFs across the data set we see that each sponge species retains a large fraction of unique GCFs (65% of all sponge symbiont GCFs). Furthermore, there is also significant individuality between the two *Geodia barretti* sample groups, i.e., samples from two different geographical locations (Norway and Canada), with unique GCFs outnumbering shared GCFs in each of the geographical locations. Nevertheless, there is also an indication of specialized metabolite conservation across the symbiotic communities of these sponge species, with all sample groups (treating *G. barretti* from Norway and Canada as separate sample groups) sharing a core of 2% (16 GCFs) of all their encoded GCFs (Fig. 2). A total of 58 GCFs were conserved across the three sponge species (merging the two *G. barretti* groups), and we recorded that an additional 200 GCFs were shared between at least two of the three sponge species. In the sponge holobiont, functional redundancy has previously been identified for primary metabolism, with nutritionally specialized guilds that span several taxonomic affiliations (48). With regard to specialized metabolism, conservation across species has been shown with the sponge ubiquitous polyketide (SUP) cluster (49), the sponge widespread fatty acid synthase (swf) cluster (50), and the sponge derived RiPP proteusins (srp) (15). Furthermore, Mohanty et al. (51). recently reported that the presence of bromotyrosine alkaloids, signature NPs that are present across phylogenetically distant sponges, is not dependent on the sponge microbiome

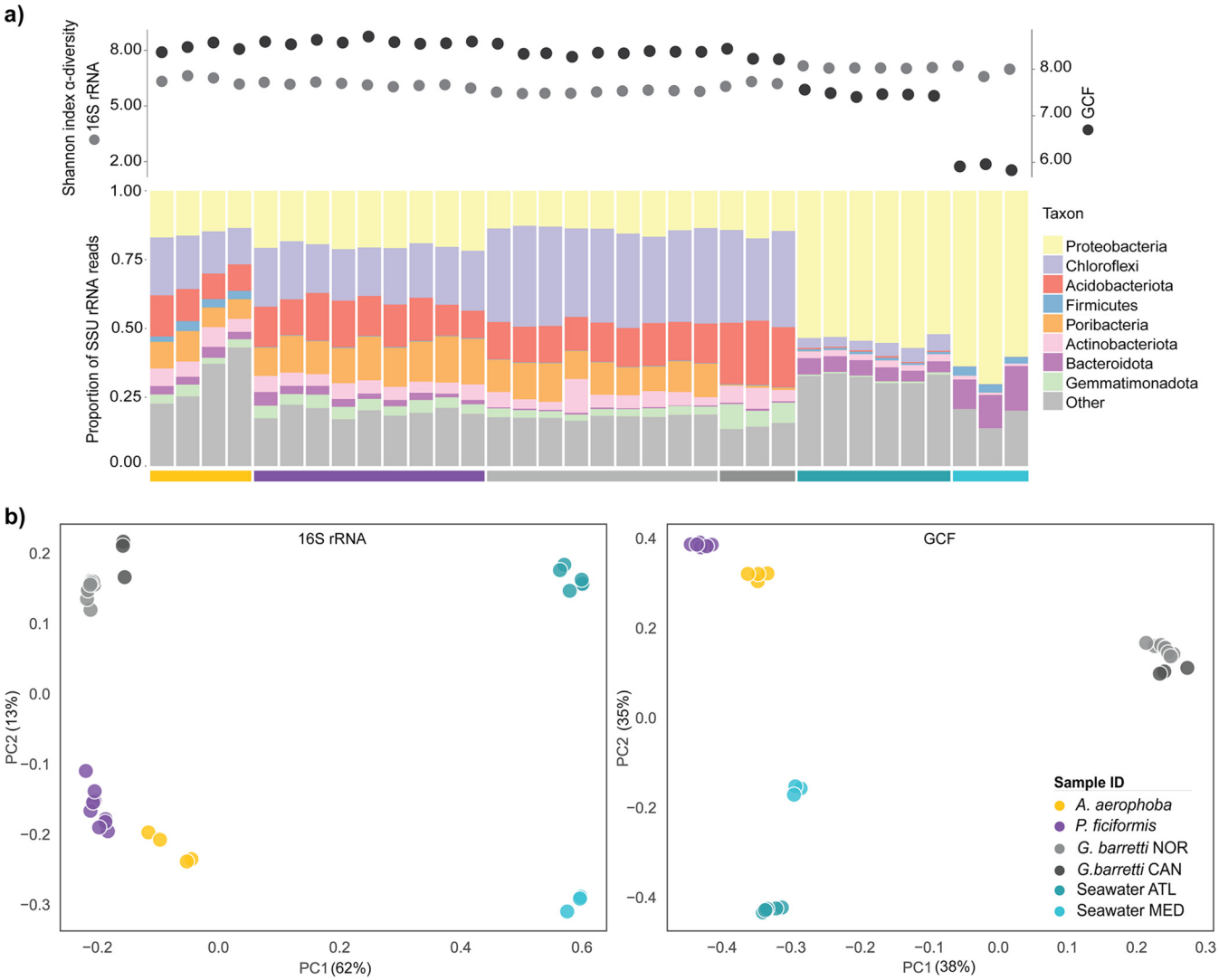

FIG 3 Biosynthetic potential and taxonomic diversity comparison of marine sponges. (a) Sample Shannon index alpha-diversity scores based on 16S SSU rRNA genus level NTU (left axis, light gray) and GCF (right axis, dark gray) content, and sample prokaryotic taxonomic profilea classified at the phylum level. (b) Beta-diversity scores for sponge samples based on 16S SSU rRNA genus level NTU (left) and GCF (right) content. The percent explained variance is denoted on each axis label.

taxonomic architecture. Although the GCF core mentioned above is mainly composed of terpene GCFs, it also includes NRPS, RiPP, and PKS GCFs. We also recovered both SUP gene clusters (49), *swf*-like gene clusters (50) (see Fig. S1 in the supplemental material), and NRPS-like ether-lipid related GCFs (discussed below [see Fig. 5]) from all sponge species, as well as RiPP srp-like clusters from two of the three species (discussed below [see Fig. 6]).

Taxonomic diversity and biosynthetic gene cluster family diversity follow similar trends. As it remains unclear from which taxa in the holobiont the predicted BGCs and GCFs derive, we aimed to gain insight into the relations between functional gene content and taxonomic diversity. Observed prokaryotic community composition (Fig. 3a) follows the commonly observed profiles (52, 53), with *Proteobacteria*, *Chloroflexi*, *Acidobacteria*, and *Poribacteria* occupying the largest fractions of these sponge-associated bacterial communities.

GCF abundance-based Shannon alpha-diversity scores do not follow taxonomic 16S SSU rRNA genus level-based alpha-diversity scores across the data set (Fig. 3a), indicating that a higher diversity of bacterial genera does not automatically lead to a more diverse biosynthetic profile (Spearman's $r = -0.4$, $P = 0.01$). Furthermore, we observe that there

are significant pairwise differences (Kruskal-Wallis, $P < 0.05$) between all sponge species' taxonomy-based alpha-diversity scores, as well as GCF-based alpha-diversity scores, with *Aplysina aerophoba* showing the highest mean taxonomy-based alpha-diversity and *Petrosia ficiformis* showing the highest mean GCF-based alpha-diversity (see Table S2). Despite the significant differences in taxonomy-based alpha-diversity, GCF abundance-based alpha-diversity did not show significant differences for the *G. barretti* NOR/CAN pair (see Table S2). This is again an indication that not all bacterial genera are equally talented, since higher taxonomic diversity (*G. barretti* CAN) does not automatically lead to higher biosynthetic diversity (*G. barretti* NOR; Table S2).

With respect to the beta-diversity of prokaryotic community composition and GCF composition, we observed that all sample groups are different, with taxonomy and GCF content generating similar patterns (Fig. 3b). The shallow Mediterranean species, *A. aerophoba* and *P. ficiformis*, have relatively similar prokaryotic communities and GCF distributions that are different from those of the deep Atlantic sponge *G. barretti* from Norway and Canada. In addition, the prokaryotic communities and GCF distribution in seawater were different for Atlantic and Mediterranean seawater and also different than those of the three sponge species. PERMANOVA (permutational multivariate analysis of variance) testing based on 16S rRNA gene genus-level nearest taxonomic unit (NTU) and GCF content revealed that all sample groups are significantly different ($P = 0.001$) in both contexts (see Table S2). These results further support the notion that, despite the presence of a shared GCF core, each sponge species has a distinct encoded specialized metabolism profile.

***Acidobacteriota* stand out as potential superproducers.** We recovered a total of 316 dereplicated MAGs, all classified at least at the phylum level (GTDB r95), from the sponge and seawater metagenomes (Fig. 4; see also Fig. S2). We observe a relatively low sharedness of MAGs, with only 3% of MAGs being found in more than one sponge species (a MAG is considered shared between samples when MAGs from different samples are clustered by dRep). When the MAGs recovered here are shared, it is mostly so within the same sponge species (*G. barretti* NOR and CAN) or between sponges sharing similar habitats (*A. aerophoba* and *P. ficiformis*). The shared MAGs are widely distributed throughout several phyla (Fig. 4), with *Acidobacteriota* MAGs showing the highest sharedness across species (30% shared MAGs in *A. aerophoba* and *P. ficiformis*). This is in line with the recent work by Robbins et al. (35), who phylogenetically characterized 1,200 MAGs derived from 30 sponge species and reported similar patterns of shared versus exclusive MAGs, with the presence of taxa that are unique to their host sponge species, as well as populations of *Acidobacteriota* that are shared across sponge species. In the specific case of the two *G. barretti* sample groups, we observe that 19% of the MAGs are shared, with proteobacterial MAGs making up the largest fraction (31%) of a set of shared MAGs that nevertheless shows diverse taxonomic assignments (Fig. 4; see also Fig. S2).

Of the 266 MAGs recovered from sponge samples, 96% contained GCFs, and of the 58 MAGs recovered from seawater samples, 83% contained GCFs. In addition, we observe an average of 3.5 GCFs per sponge MAG and 2 GCFs per seawater MAG. *Acidobacteriota* have been known as talented NP producers (42, 54) and also emerge here as a functionally diverse phylum displaying unparalleled biosynthetic potential, represented by 28 MAGs with an average of 6.6 GCFs per MAG. Another diverse and consistently talented phylum is *Latescibacterota* (average 4.4 GCFs per MAG), comprising well-known members of sponge holobiont communities (30, 35, 36) that have only recently been linked to NP production (15) and thus show particular interest for discovery. We observe that members of both these phyla are present in all three sponge species. In addition, two other prolific phyla, yet represented by fewer MAGS, are *Nitrospirota* (average, 8.3 GCFs per MAG) and *Desulfobacterota_B* (average, 6.0 GCFs per MAG). The candidate phylum *Tectomicrobia* is also represented here by three MAGs, respectively encoding 2, 5, and 7 GCFs, and were classified as genus "SXND01" within the family *Entotheonellaceae*. "*Ca.* Tectomicrobia" is famous for its biosynthetically prolific candidate genus *Entotheonella* (25), with other lineages within "*Ca.*

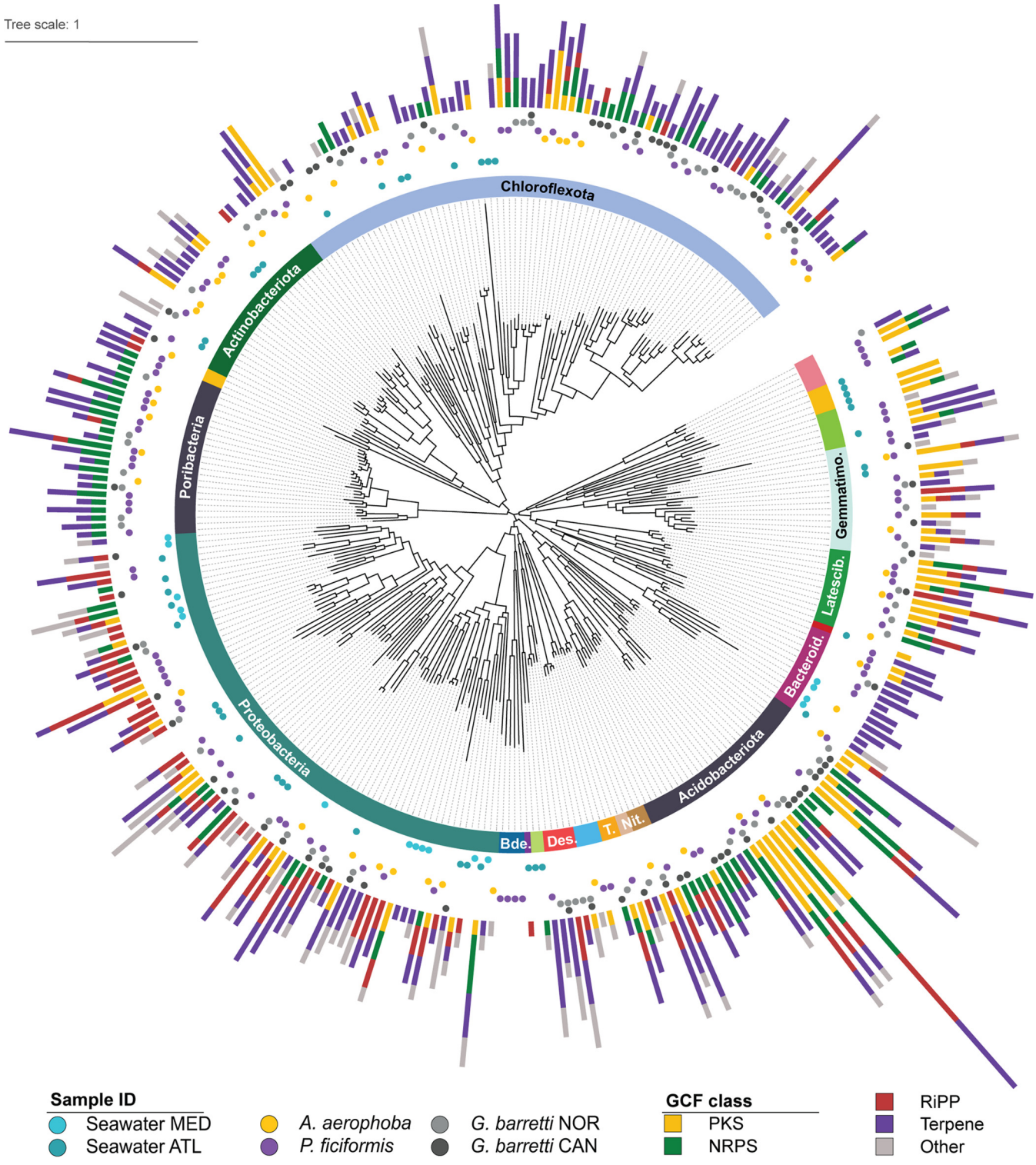

FIG 4 GCFs present in the recovered MAGs. Cladogram based on GTDB classification of MAGs constructed in this study. Layer 1: phylum classification (truncated phyla are *Gemmatimonadota*, *Latescibacterota*, *Bacteroidota*, *Nitrospinota*, *Nitrospinota_A Tectomicrobia*, *Desulfobacterota_B*, and *Bdellovibrionota*). The complete MAG classification is shown in Fig. S2. Layer 2: sample groups where MAGs are observed. Layer 3: GCF content colored by class.

Tectomicrobia" currently described as biosynthetically poor (E. Peters, unpublished data). Finally, approximately 40% of all GCFs remained unbinned. This is likely a result of the ineffectivity of current MAG binning methods to successfully process plasmids, mobile elements and genomic islands (55), since BGCs are often located in such mobile regions (56–59).

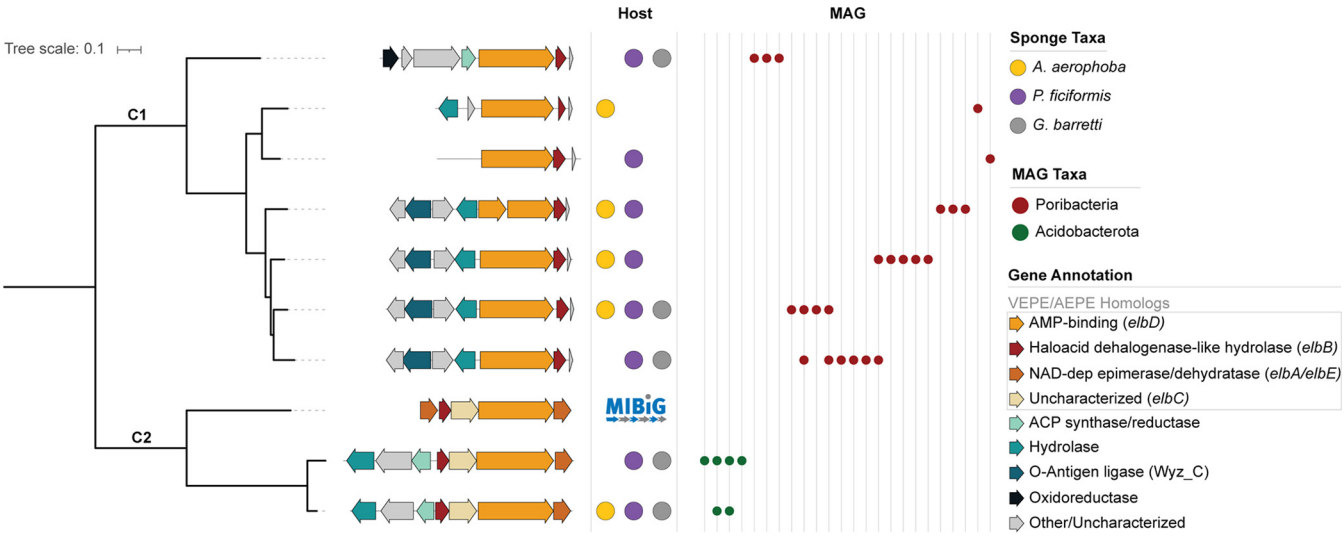

FIG 5 Distribution and characterization of the VEPE-related GCFs. The FastTree phylogeny of VEPE-associated GCFs, each characterized by a representative BGC, based on sequence similarity of the A-domain (AMP-binding), is shown. A detailed overview of these GCFs is provided in Fig. S3. C1 and C2 refer to clades 1 and 2. Genes are colored based on predicted function. Each BGC is annotated by its presence in host sponge species and in MAGs.

The previously mentioned SUP (49) and *swf*-like (50) GCFs that were obtained from all sponge species appear to be independent of bacterial taxonomy. They were found in several MAGs of the phyla *Chloroflexota*, *Spirochaetota*, *Proteobacteria*, and *Acidobacteriota* in the case of SUP-like GCFs, and *Nitrospirota*, *Latescibacterota*, and *Acidobacteriota* in the case of *swf*-like GCFs. Multiple MAGs encoding SUP and *swf*-like GCFs belonging to different phyla were observed within the same sponge holobiont. However, despite the similarity of taxa harboring the SUP-like and *swf*-like GCFs at the phylum level, at the species level (<95% gANI) the majority of these MAGs are specific to a single sponge species (see Table S4).

Widespread ether-lipid-associated BGCs in sponges. Within the shared GCF core, we identified nine nonribosomal peptide synthetase (NRPS)-like GCFs with a similar architecture: an NRPS-like core gene (*elbD* homolog) containing fatty acyl coenzyme A (acyl-CoA)-like reductase, acyl-CoA synthetase, thiolation, and acylglycerolphosphate acyltransferase domains. The NRPS-like core gene was consistently flanked by hydrolases/dehydratases, as well as oxidoreductases and genes involved in fatty acid biosynthesis (Fig. 5). These GCFs show similarity to a known BGC encoding the production of vinyl-/alkyl-ether lipids (VEPE/AEPE; MiBIG accession no. BGC0000871). The VEPE/AEPE lipids are produced by *Myxococcus xanthus* DK 1622 as extracellular signals guiding fruiting body morphogenesis and sporulation (46, 60). It is postulated that these lipids are generated via modification of phospholipids originating from the cell membrane with participation of BGC0000871's genes *elbB*, *elbD*, and *elbE*, as well as the additional desaturase *carF* (60, 61). An *elbD* homolog of poribacterial origin has also been identified by Lorenzen et al. (46).

We observe architectural changes in the adjacent genes of the cluster across the set that are congruent with the clades generated based on sequence similarity of the A domain (Fig. 5). Additional analysis of adenylation domain active site specificity-conferring residues (AdenylPred [62]; see Table S5) indicates potentially different functional classes and substrate specificities for the two clades, which suggests diversity in the chemical compounds produced by the encoded machinery. Gene clusters from the two clades are associated with distinct bacterial taxa, being specific to *Acidobacteriota* and *Poribacteria* (GTDB r95) MAGs, with any given MAG harboring a maximum of two of the nine GCFs (see Table S6).

Ether lipids seen in the sponge holobiont are often linked to pathogen defense by showing antimicrobial activity (63–66). However, the biosynthetic origin and pathway for these molecules is currently undescribed. Ultimately experimental work will be necessary to determine the function and biosynthesis of these ether lipids within the sponge holobiont.

High RiPP diversity in sponge bacterial symbionts. Even though the sponge holobiont is recognized as an extensive source of diverse NPs, its inventory of ribosomally synthesized and posttranslationally modified peptides (RiPPs) has remained largely undescribed, with one exception being the proteusin polytheonamides (15, 25, 67). Here, we expand the known repertoire in sponge holobionts and showcase sponges' RiPP diversity by identifying 17 uncharacterized RiPP families which seem to be widespread in sponge holobionts. In some of these cases, GCFs are recovered from all sponge species, which points toward an important role of the produced NPs in the context of sponge-microbe symbiosis (Fig. 6).

RiPPs constitute a class of natural products that is produced from a ribosomally synthesized precursor peptide, composed of an N-terminal leader and a C-terminal core, which is modified by biosynthetic enzymes encoded in the BGC, and matured in a final proteolytic cleavage step (68). We see a high incidence of RiPP BGCs with predicted nitrile hydratase-like leader peptide (NHLP) domains, which have been identified as proteusin precursor peptides, as well as nif11-like leader peptide domains associated with thiazole/oxazole-modified microcins (TOMM) precursor peptides (69). Encoded within these BGCs are a number of modifying enzymes that span several RiPP subclasses: YcaO cyclodehydratases (70, 71), radical *S*-adenosylmethionine (rSAM) enzymes (72) including PoyD (67), lanthionine synthetase C-like (LanC) enzymes (73), nitroreductases (NitroR) (74), and TfuA-like enzymes (75). Other modifying enzymes include oxidoreductases, group transferases, and halogenases. Several RiPP subclasses contain YcaO-generated azol(in)e heterocycles (71), such as the linear azol(in)e-containing peptides (LAP) and thiopeptides, and are often linked to pharmacologically interesting bioactivities (76–78). The potential for azol(in)e-containing RiPPs has been found widespread in bacterial genomes (79) but has only recently been reported in marine sponge microbiomes with the identification of the *srp* clusters by Nguyen et al. (15). We predict an additional structurally diversified set of azol(in)e-containing RiPPs to this repertoire, expanding the combinatorial space of modifying enzymes and hence the potential for production of bioactive NPs.

In addition, we observed four GFCs that harbor a cupin-4 modifying enzyme in addition to an NHLP domain, which has also been connected to proteusin biosynthesis (80). The cupin superfamily is functionally highly diverse, with reported hydroxylation, epoxidation, dioxygenation, decarboxylation, dehydration, and halogenation activities (81–83). Although these genes appear in marine environments, they are not yet regarded as a typical constituent of the sponge microbiome biosynthetic array (84, 85). We further identified several GCFs predicted to be responsible for the production of highly modified lassopeptides, encoding additional hydrolases, oxidoreductases, and group transferases (sulfo-, glycol-, and nucletotidyl-). These include GCFs that are shared among all sponge species. Lassopeptides have been found in some sponge-associated microbiomes (86, 87), but their presence together with diverse flanking enzyme-encoding genes has not been previously reported in this environmental niche. While the functions of these BGCs and their metabolic products remains elusive, this data set expands and diversifies the set of sponge-holobiont-encoded RiPPs and predicts sponge microbiomes to be prolific sources of RiPP NPs.

Conclusion. Here, we describe a systematic analysis of the specialized metabolite diversity landscape of several marine sponge holobionts, whose metagenomes contain a large number of uncharacterized BGCs. We show that while there is high biosynthetic diversity that is unique to each sponge species' symbiotic community, a small functional core is conserved throughout the studied sponge species. This functional core includes a novel group of NRPS-like ether lipid-associated GCFs widespread across sponges and a number of RiPPs. We recovered a consistent set of biosynthetically potent MAGs from the metagenomes, with *Acidobacteriota* and *Latescibacterota* standing out as potential prolific NP producers. This study thus contributes to a growing body of research uncovering the network of NPs with elusive functions that are generated by marine sponge microbiomes.

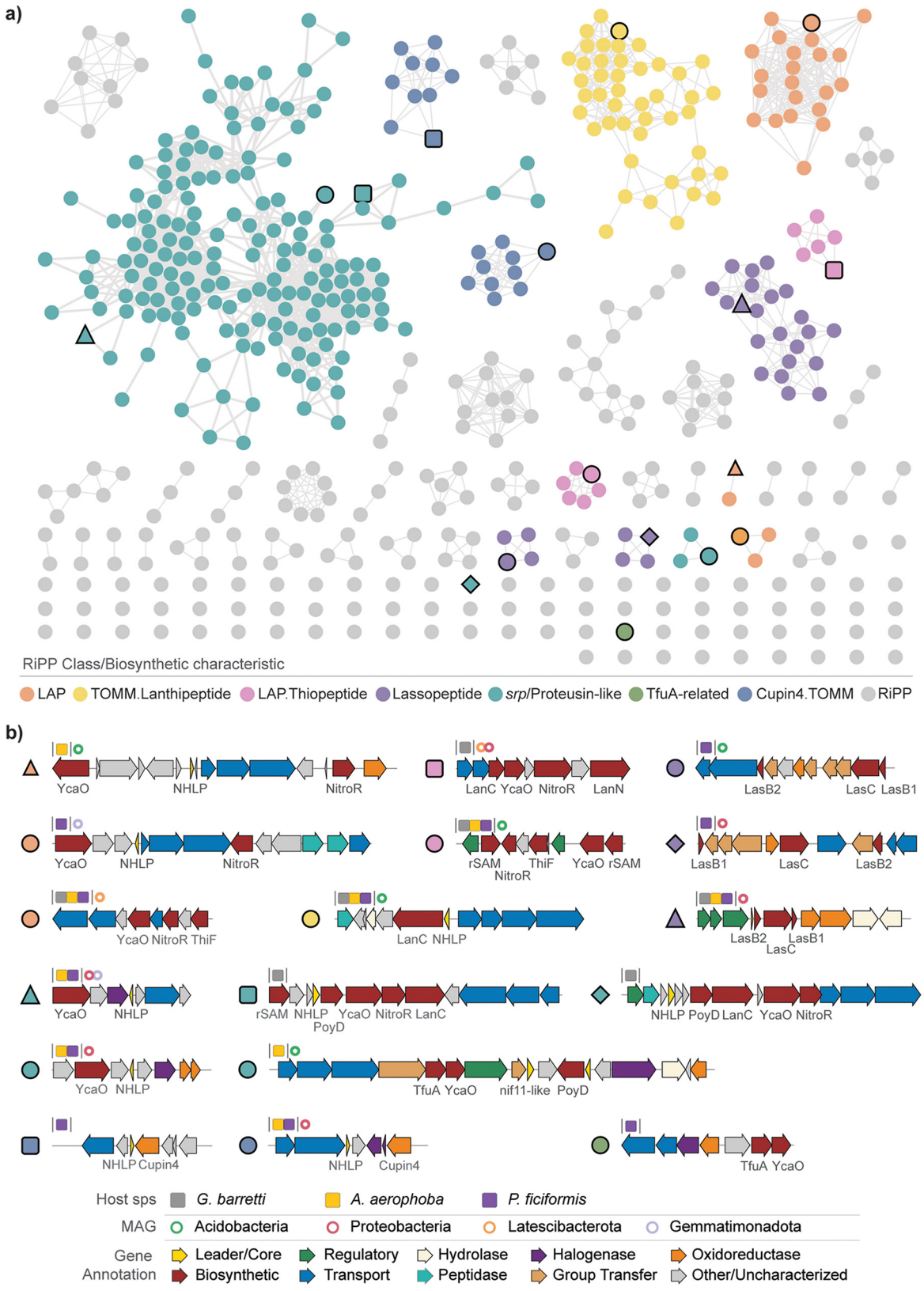

FIG 6 RiPP diversity. (a) RiPP BGCs represented in a BiG-SCAPE-generated similarity network. Highlighted nodes are colored by biosynthetic class/characteristic, boldface node shapes link to the specific BGCs described below, and gray nodes represent nonhighlighted RiPP BCGs. (b) Representative BGCs from each highlighted GCF are depicted and characterized by the predicted class, the presence of GCF in a host sponge, and MAG (GTDB r95). Genes are colored by predicted function.

MATERIALS AND METHODS

Sponge collection. Norway *Geodia barretti* (*G. barretti* NOR) samples (gb1-gb10) were collected and processed by E. Peters et al. (unpublished data), and Canada *Geodia barretti* (*G. barretti* CAN) samples (gb126, gb278, gb305) were collected and processed by Steffen et al. (107). Atlantic Seawater (Seawater ATL) samples (gb1_f – gb10_f) were collected onboard R/V Hans Brattström of the University of Bergen from Korsfjord, Bergen, Norway (60°8.13′N, 5°6.7′E) in September and October 2017 by filtering 2 L of seawater through polyvinylidene difluoride membrane filters (pore size, 0.22 μm; diameter, 47 mm; Merck Millipore, Burlington, MA). Filters were snap-frozen in liquid nitrogen and stored at −80°C. *Aplysina aerophoba* and Mediterranean Seawater (Seawater_MED) samples were collected and processed as described by Chaib De Mares et al. (88) and are publicly available (88). *Petrosia ficiformis* sampling took place in August 2018 at a semisubmerged marine cave (5- to 6-m depth) with internal freshwater springs in Sfakia, Greece (35°12′N, 24°7′E) in a collaborative effort with the Hellenic Centre for Marine Research (HCMR). Immediately after collection, the samples were transferred to liquid nitrogen and stored at −80°C.

Total DNA extraction and metagenomic sequencing. *A. aerophoba*, *G. barretti*, and *P. ficiformis* sponge samples were crushed in liquid nitrogen to a fine power with pestle and mortar. A 200-mg portion of sponge tissue powder was further disrupted by bead beating using milling balls (5 × 2 mm + 2 × 5 mm), followed by two steps of shaking for 20 s at 4,000 rpm in a Precellys 24 tissue homogenizer (Bertin Instruments, Montigny-le-Bretonneux, France) (89). Tissue lysate was further used for DNA extraction with an AllPrep DNA/RNA/protein minikit (Qiagen, Hilden, Germany). Total DNA extraction of seawater filters was done using the entire filter, following the protocol described above. DNA extracted from water filter replicates gb5_f and gb6_f was pooled prior to sequencing to meet minimum DNA quantity requirements. The extracted DNA was further cleaned by a collagenase treatment using C9891 collagenase from *Clostridium histolyticum* (Sigma-Aldrich, St. Louis, MO) at a concentration of 2.5 mg/mL for 30 min at 4°C with vortexing every 5 min at max speed for 10 s. It was then purified using a MasterPure Gram-positive DNA purification kit (Lucigen, Middleton, WI) according to the manufacturers' instructions and passed through Illustra MicroSpin S-400 HR columns (GE Healthcare, Chicago, IL). *G. barretti*, *P. ficiformis*, and Seawater_ATL total DNA was obtained; this was sequenced by Novogene (Hong Kong, China) using an Illumina HiSeq PE150 platform. *A. aerophoba* and Seawater_MED total DNA was sequenced by Research Group Genome Analytics (GMAK) at the DSMZ (Braunschweig, Germany) using Illumina HiSeq PE100. *A. aerophoba* DNA was also prepared for Pacific Biosciences (PacBio, Menlo Park, CA) long-read sequencing and was processed as described previously (90) and sequenced at the DSMZ. Complete sample metadata can be found in Table S1 in the supplemental material.

Quality trimming and adapter removal. Illumina HiSeq read adapter removal, quality filtering and normalization was done using the BBduk.sh script from the BBTools suite v37.64 (91), following user guide instructions, with the parameters ktrim=r k=23 mink=7 hdist=1 tpe tbo qtrim=rl trimq=20 ftm=5 maq=20. The minlen parameter was set to 30 for *A. aerophoba* and Med_SW samples, and to 50 for all remaining samples. BBDuk (91) was also used to remove sequencing artifacts and phi X contamination, with default settings.

Metagenomic assembly. Reads were normalized for coverage with BBNorm (91) with the parameters target=100 min=5 for *Petrosia ficiformis* (sequenced and processed at a later stage), and target=200 min=3 for all other samples. As SPAdes v3.12 (92) hybrid mode (–pacbio) does not support coassembly, *A. aerophoba* sample Aply22 sequencing replicates were merged prior to hybrid assembly. *A. aerophoba* filtered Illumina HiSeq reads and PacBio reads were assembled with SPAdes v3.12 (92) using the –meta and –only-assembler flags. Filtered Illumina HiSeq reads from all other samples were assembled with SPAdes v3.12 (92) using the –meta and –only-assembler flags.

MAG binning, dereplication, and classification. Contigs were binned using metaWRAP v1.2 (93) with minimum completeness of 75% and maximum contamination of 10% (see Table S3), using MaxBin2 (94), metaBAT2 (95), and CONCOCT (96). The obtained bins were dereplicated using dRep (97) v2.5.4 with default parameters for primary clustering and secondary clustering using parameters –S_algorithm gANI –S_ani 0.95 (98). The dereplicated bins were taxonomically classified using the GTDB-Toolkit (99) v1.1.0 (GTDB-Tk) classify workflow. A phylogenetic tree of the dereplicated bins was created using the multiple sequence alignment generated in the GTDB-Tk (99) align workflow with FastTree (100) v2.1.11, default parameters. This tree was visualized and annotated using the Interactive Tree of Life (101) (iTOL v6) online tool.

Metagenome 16S rRNA gene taxonomic classification. 16S rRNA gene sequences were extracted and characterized using Phyloflash (102) v3 with the SILVA rRNA gene database (103) v38.1, the parameters -taxlevel 6 -poscov and -readlength 150 for *G. barretti*, Seawater_Atl and *P. ficiformis* samples, 100 for *A. aerophoba* and Seawater_Med Taxa relative abundance plots (–task barplot –level 6), and an NTU table (–task ntu_table –level 6) were generated with phyloFlash_compare.pl.

Biosynthetic gene cluster analysis. BGC prediction was performed for all contigs over 4000 bp in length using antiSMASH (40) v5 using the following parameters: –cb-general –cb-subclusters –cb-knownclusters –minlength 4000 –hmmdetection-strictness relaxed –genefinding-tool prodigal-m –clusterhmmer –asf –smcog-trees –pfam2go. BiG-SCAPE (41) v1.0.1 was run on all predicted BGCs using parameters –mix -v –mode auto –mibig –cutoffs 0.5 –include_singletons. BiG-SCAPE (41) network files were processed by in-house Python scripts available at https://github.com/CatarinaCarolina/sponge_meta_BGC to generate Fig. 1a, b, and e, as well as Fig. 5, using the Python package UpSetPlot v0.4.1. Phylogenetic analysis of NRPS-like GCFs was conducted using CORASON (41) v1, default parameters, with gb8_2 contig 859 gene 9 as a query and MIBiG BGC0000871.1 as reference BGC, and nine GCFs were selected for further analysis (see Fig. S3). A representative BGC was selected from each of these, based on BGC completeness and best representation of adjacent gene diversity encoded in the GCF. The respective AMP domain amino acid sequences were used in a multiple sequence alignment (MSA) done with

Muscle (104) v3.8.31. FastTree (100) v2.1.11 processed the MSA into a phylogenetic tree, visualized with iTOL (101). A RiPP-specific BiG-SCAPE (41) run was carried out by selecting all RiPP BGCs previously identified, excluding those classified with the antiSMASH (40) rules DUF692 and TIGR03975, since these BGCs were likely false positives based on manual inspection. GCF abundance, i.e., normalized RPKM (reads per kilobase per million), was calculated using BiG-MAP (105) with parameters -tg 0 -c 0.5 for the family module, and otherwise at default settings.

Sequence estimated coverage. Sample read redundancy estimation was calculated using Nonpareil (106) v3.304 with parameters -T kmer -X 1000000. Nonpareil curves were built using the R package Nonpareil in RStudio, R v4.0.3.

Statistical and diversity analysis. The Phyloflash (101) bacterial/16S-rRNA-gene-derived NTU table, normalized by relative abundance, and the BiG-MAP (105) normalized RPKM table were used to calculate Sample Shannon alpha-diversity scores with the Python package skbio.diversity.alpha.shannon, as well as to generate a Bray-Curtis dissimilarity matrix with the Python package scipy.spatial.distance.braycurtis. Pearson's correlation score for Shannon alpha-diversity scores was calculated using the Python package scipy.stats.pearsonr. A principal coordinate analysis using these same distance matrices was performed using the Python package skbio.stats.ordination.pcoa. ERMANOVA was calculated for both distance matrices grouping by sample type using the Python package skbio.stats.distance.permanova. Python scripts used to carry out this analysis are available at https://github.com/CatarinaCarolina/sponge_meta_BGC.

MAG and GCF data integration. MAG iTOL (101) annotation tables were generating by processing GTDB-Tk and BiG-SCAPE outputs with in-house Python scripts available at https://github.com/CatarinaCarolina/sponge_meta_BGC.

Data availability. The data for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB51534. Python scripts created for this analysis are available (https://github.com/CatarinaCarolina/sponge_meta_BGC).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 2.8 MB.

FIG S2, PDF file, 1.2 MB.

FIG S3, PDF file, 0.7 MB.

TABLE S1, PDF file, 0.2 MB.

TABLE S2, PDF file, 0.2 MB.

TABLE S3, PDF file, 0.1 MB.

TABLE S4, PDF file, 0.2 MB.

TABLE S5, PDF file, 0.3 MB.

TABLE S6, PDF file, 0.5 MB.

ACKNOWLEDGMENTS

We thank Ellen Kenchington for providing the Canadian *Geodia barretti* samples, the late Hans Tore Rapp for help in sampling the Norwegian *G. barretti* samples, and Vasilis Gerovasileiou and the HCMR (Hellenic Centre for Marine Research) for the collaboration in sampling the *P. ficiformis* samples. We thank Vittorio Tracanna, Joris Louwen, and Michelle Schorn for valuable advice and discussions.

This research was financially supported by the VLAG NWO Ph.D. project Mare Incognita to C.L., the European Commission through Horizon2020 project SponGES (grant agreement 679849) to D.S. and A.G. (Asimenia), and the BluePharmTrain project (grant agreement 607786) to D.S.

This project has also received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement no. 101000392 (MARBLES). This output reflects only the author's view and the Research Executive Agency (REA) cannot be held responsible for any use that may be made of the information contained therein.

C.L. provided research design, *G. barretti* and and Seawater_ATL eDNA extraction, and seawater, *G. barretti*, and *A. aerophoba* sequence QC filtering and assembly. C.L. performed all sample BGC prediction and characterization, binning, taxonomy profile prediction, Nonpareil estimates, and downstream analysis. C.L. also wrote the manuscript. An.G. performed *P. ficiformis* sampling, eDNA extraction, QC filtering, and assembly. As.G. performed *G. barretti* and Seawater_ATL sampling and eDNA extraction. M.C.D.M. performed SW_Med and *A. aerophoba* sampling and eDNA extraction. J.V.D.O. performed research design and manuscript reviewing/editing. M.H.D. performed research design and

manuscript reviewing/editing. D.S. performed research design, manuscript reviewing/editing, and *G. barretti* and Seawater_ATL sampling.

M.H.M. is a cofounder of Design Pharmaceuticals and a member of the scientific advisory board of Hexagon Bio.

REFERENCES

1. 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.
2. 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.
3. Paul VJ, Freeman CJ, Agarwal V. 2019. Chemical ecology of marine sponges: new opportunities through "-omics." Integr Comp Biol 59:765–776. https://doi.org/10.1093/icb/icz014.
4. Freeman CJ, Thacker RW, Baker DM, Fogel ML. 2013. Quality or quantity: is nutrient transfer driven more by symbiont identity and productivity than by symbiont abundance? ISME J 7:1116–1125. https://doi.org/10.1038/ismej.2013.7.
5. Webster NS, Thomas T. 2016. The sponge hologenome. mBio 7:e00135-16. https://doi.org/10.1128/mBio.00135-16.
6. 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.
7. Pita L, Rix L, Slaby BM, Franke A, Hentschel U. 2018. The sponge holobiont in a changing ocean: from microbes to ecosystems. Microbiome 6:46. https://doi.org/10.1186/s40168-018-0428-1.
8. Webster NS, Taylor MW. 2012. Marine sponges and their microbial symbionts: love and other relationships. Environ Microbiol 14:335–346. https://doi.org/10.1111/j.1462-2920.2011.02460.x.
9. Hentschel U, Fieseler L, Wehrl M, Gernert C, Steinert M, Hacker J, Horn M. 2003. Microbial diversity of marine sponges. Prog Mol Subcell Biol 37:59–88. https://doi.org/10.1007/978-3-642-55519-0_3.
10. Gloeckner V, Wehrl M, Moitinho-Silva L, Gernert C, Schupp P, Pawlik JR, Lindquist NL, Erpenbeck D, Wörheide G, Hentschel U. 2014. The HMA-LMA dichotomy revisited: an electron microscopical survey of 56 sponge species. Biol Bull 227:78–88. https://doi.org/10.1086/BBLv227n1p78.
11. Moitinho-Silva L, Steinert G, Nielsen S, Hardoim CCP, Wu Y-C, McCormack GP, López-Legentil S, Marchant R, Webster N, Thomas T, Hentschel U. 2017. Predicting the HMA-LMA status in marine sponges by machine learning. Front Microbiol 8:752. https://doi.org/10.3389/fmicb.2017.00752.
12. Easson CG, Thacker RW. 2014. Phylogenetic signal in the community structure of host-specific microbiomes of tropical marine sponges. Front Microbiol 5:532–511. https://doi.org/10.3389/fmicb.2014.00532.
13. 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.
14. 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.
15. Nguyen NA, Lin Z, Mohanty I, Garg N, Schmidt EW, Agarwal V. 2021. An obligate peptidyl brominase underlies the discovery of highly distributed biosynthetic gene clusters in marine sponge microbiomes. J Am Chem Soc 143:10221–10231. https://doi.org/10.1021/jacs.1c03474.
16. Morita M, Schmidt EW. 2018. Parallel lives of symbionts and hosts: chemical mutualism in marine animals. Nat Prod Rep 35:357–378. https://doi.org/10.1039/c7np00053g.
17. Rix L, Ribes M, Coma R, Jahn MT, de Goeij JM, van Oevelen D, Escrig S, Meibom A, Hentschel U. 2020. Heterotrophy in the earliest gut: a single-cell view of heterotrophic carbon and nitrogen assimilation in sponge-microbe symbioses. ISME J 14:2554–2567. https://doi.org/10.1038/s41396-020-0706-3.
18. Pawlik JR. 2011. The chemical ecology of sponges on Caribbean reefs: natural products shape natural systems. Bioscience 61:888–898. https://doi.org/10.1525/bio.2011.61.11.8.
19. Smith TE, Pond CD, Pierce E, Harmer ZP, Kwan J, Zachariah MM, Harper MK, Wyche TP, Matainaho TK, Bugni TS, Barrows LR, Ireland CM, Schmidt EW. 2018. Accessing chemical diversity from the uncultivated symbionts of small marine animals. Nat Chem Biol 14:179–185. https://doi.org/10.1038/nchembio.2537.
20. Crawford JM, Clardy J. 2011. Bacterial symbionts and natural products. Chem Commun (Camb) 47:7559–7566. https://doi.org/10.1039/c1cc11574j.
21. Song H, Hewitt OH, Degnan SM. 2021. Arginine biosynthesis by a bacterial symbiont enables nitric oxide production and facilitates larval settlement in the marine-sponge host. Curr Biol 31:433–437. https://doi.org/10.1016/j.cub.2020.10.051.
22. Moitinho-Silva L, Díez-Vives C, Batani G, Esteves AI, Jahn MT, Thomas T. 2017. Integrated metabolism in sponge-microbe symbiosis revealed by genome-centered metatranscriptomics. ISME J 11:1651–1666. https://doi.org/10.1038/ismej.2017.25.
23. Lopera J, Miller IJ, McPhail KL, Kwan JC. 2017. Increased biosynthetic gene dosage in a genome-reduced defensive bacterial symbiont. mSystems 2:e00096-17. https://doi.org/10.1128/mSystems.00096-17.
24. Nakabachi A, Ueoka R, Oshima K, Teta R, Mangoni A, Gurgui M, Oldham NJ, van Echten-Deckert G, Okamura K, Yamamoto K, Inoue H, Ohkuma M, Hongoh Y, Miyagishima S-y, Hattori M, Piel J, Fukatsu T. 2013. Defensive bacteriome symbiont with a drastically reduced genome. Curr Biol 23:1478–1484. https://doi.org/10.1016/j.cub.2013.06.027.
25. Rust M, Helfrich EJN, Freeman MF, Nanudorn P, Field CM, Rückert C, Kündig T, Page MJ, Webb VL, Kalinowski J, Sunagawa S, Piel J. 2020. A multiproducer microbiome generates chemical diversity in the marine sponge *Mycale hentscheli*. Proc Natl Acad Sci U S A 117:9508–9518. https://doi.org/10.1073/pnas.1919245117.
26. Calcabrini C, Catanzaro E, Bishayee A, Turrini E, Fimognari C. 2017. Marine sponge natural products with anticancer potential: an updated review. Mar Drugs 15:310. https://doi.org/10.3390/md15100310.
27. Carroll AR, Copp BR, Davis RA, Keyzers RA, Prinsep MR. 2019. Marine natural products. Nat Prod Rep 36:122–173. https://doi.org/10.1039/C8NP00092A.
28. Nayfach S, Rodriguez-Mueller B, Garud N, Pollard KS. 2016. An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. Genome Res 26:1612–1625. https://doi.org/10.1101/gr.201863.115.
29. Pachiadaki MG, Brown JM, Brown J, Bezuidt O, Berube PM, Biller SJ, Poulton NJ, Burkart MD, La Clair JJ, Chisholm SW, Stepanauskas R. 2019. Charting the complexity of the marine microbiome through single-cell genomics. Cell 179:1623–1635. https://doi.org/10.1016/j.cell.2019.11.017.
30. Storey MA, Andreassend SK, Bracegirdle J, Brown A, Keyzers RA, Ackerley DF, Northcote PT, Owen JG. 2020. Metagenomic exploration of the marine sponge *Mycale hentscheli* uncovers multiple polyketide-producing bacterial symbionts. mBio 11:e02997-19. https://doi.org/10.1128/mBio.02997-19.
31. Tianero MD, Balaich JN, Donia MS. 2019. Localized production of defence chemicals by intracellular symbionts of *Haliclona* sponges. Nat Microbiol 4:1149–1159. https://doi.org/10.1038/s41564-019-0415-8.
32. Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, de Bruijn I, Chooi YH, Claesen J, Coates RC, Cruz-Morales P, Duddela S, Düsterhus S, Edwards DJ, Fewer DP, Garg N, Geiger C, Gomez-Escribano JP, Greule A, Hadjithomas M, Haines AS, Helfrich EJN, et al. 2015. Minimum information about a biosynthetic gene cluster. Nat Chem Biol 11:625–631.
33. Wilson MC, Mori T, Rückert C, Uria AR, Helf MJ, Takada K, Gernert C, Steffens UAE, Heycke N, Schmitt S, Rinke C, Helfrich EJN, Brachmann AO, Gurgui C, Wakimoto T, Kracht M, Crüsemann 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.
34. Freeman MF, Helf MJ, Bhushan A, Morinaka BI, Piel J. 2017. Seven enzymes create extraordinary molecular complexity in an uncultivated bacterium. Nat Chem 9:387–395. https://doi.org/10.1038/nchem.2666.

35. Robbins SJ, Song W, Engelberts JP, Glasl B, Slaby BM, Boyd J, Marangon E, Botté ES, Laffy P, Thomas T, Webster NS. 2021. A genomic view of the microbiome of coral reef demosponges. ISME J 15:1641–1654. https://doi.org/10.1038/s41396-020-00876-9.
36. Engelberts JP, Robbins SJ, de Goeij JM, Aranda M, Bell SC, Webster NS. 2020. Characterization of a sponge microbiome using an integrative genome-centric approach. ISME J 14:1100–1110. https://doi.org/10.1038/s41396-020-0591-9.
37. Sugimoto Y, Camacho FR, Wang S, Chankhamjon P, Odabas A, Biswas A, Jeffrey PD, Donia MS. 2019. A metagenomic strategy for harnessing the chemical repertoire of the human microbiome. Science 366:eaax9176. https://doi.org/10.1126/science.aax9176.
38. Robinson SL, Piel J, Sunagawa S. 2021. A roadmap for metagenomic enzyme discovery. Nat Prod Rep 38:1994–2023.
39. Kountz DJ, Balskus EP. 2021. Leveraging microbial genomes and genomic context for chemical discovery. Acc Chem Res 54:2788–2797. https://doi.org/10.1021/acs.accounts.1c00100.
40. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. AntiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81–W87. https://doi.org/10.1093/nar/gkz310.
41. Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, Kautsar SA, Tryon JH, Parkinson EI, De Los Santos ELC, Yeong M, Cruz-Morales P, Abubucker S, Roeters A, Lokhorst W, Fernandez-Guerra A, Cappelini LTD, Goering AW, Thomson RJ, Metcalf WW, Kelleher NL, Barona-Gomez F, Medema MH. 2020. A computational framework to explore large-scale biosynthetic diversity. Nat Chem Biol 16:60–68. https://doi.org/10.1038/s41589-019-0400-9.
42. Crits-Christoph A, Diamond S, Butterfield CN, Thomas BC, Banfield JF. 2018. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. Nature 558:440–444. https://doi.org/10.1038/s41586-018-0207-y.
43. Nayfach S, Roux S, Seshadri R, Udwary D, Varghese N, Schulz F, Wu D, Paez-Espino D, Chen I-M, Huntemann M, Palaniappan K, Ladau J, Mukherjee S, Reddy TBK, Nielsen T, Kirton E, Faria JP, Edirisinghe JN, Henry CS, Jungbluth SP, Chivian D, Dehal P, Wood-Charlson EM, Arkin AP, Tringe SG, Visel A, Woyke T, Mouncey NJ, Ivanova NN, Kyrpides NC, Eloe-Fadrosh EA, IMG/M Data Consortium. 2021. A genomic catalog of Earth's microbiomes. Nat Biotechnol 39:499–509.
44. Lackner G, Peters EE, Helfrich EJN, 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.
45. Becher PG, Verschut V, Bibb MJ, Bush MJ, Molnár BP, Barane E, Al-Bassam MM, Chandra G, Song L, Challis GL, Buttner MJ, Flärdh K. 2020. Developmentally regulated volatiles geosmin and 2-methylisoborneol attract a soil arthropod to *Streptomyces* bacteria promoting spore dispersal. Nat Microbiol 5:821–829. https://doi.org/10.1038/s41564-020-0697-x.
46. Lorenzen W, Ahrendt T, Bozhüyük KAJ, Bode HB. 2014. A multifunctional enzyme is involved in bacterial ether lipid biosynthesis. Nat Chem Biol 10:425–427. https://doi.org/10.1038/nchembio.1526.
47. Kautsar SA, Blin K, Shaw S, Navarro-Muñoz JC, Terlouw BR, van der Hooft JJJ, van Santen JA, Tracanna V, Suarez Duran HG, Pascal Andreu V, Selem-Mojica N, Alanjary M, Robinson SL, Lund G, Epstein SC, Sisto AC, Charkoudian LK, Collemare J, Linington RG, Weber T, Medema MH. 2020. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. Nucleic Acids Res 48:D454–D458.
48. 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.
49. Hochmuth T, Piel J. 2009. Polyketide synthases of bacterial symbionts in sponges: evolution-based applications in natural products research. Phytochemistry 70:1841–1849. https://doi.org/10.1016/j.phytochem.2009.04.010.
50. Della Sala G, Hochmuth T, Costantino V, Teta R, Gerwick W, Gerwick L, Piel J, Mangoni A. 2013. Polyketide genes in the marine sponge *Plakortis simplex*: a new group of mono-modular type I polyketide synthases from sponge symbionts. Environ Microbiol Rep 5:809–818. https://doi.org/10.1111/1758-2229.12081.
51. Mohanty I, Tapadar S, Moore SG, Biggs JS, Freeman CJ, Gaul DA, Garg N, Agarwal V. 2021. Presence of bromotyrosine alkaloids in marine sponges is independent of metabolomic and microbiome architectures. mSystems 6:e01387-20. https://doi.org/10.1128/mSystems.01387-20.
52. Bayer K, Moitinho-Silva L, Brümmer F, Cannistraci CV, Ravasi T, Hentschel U. 2014. GeoChip-based insights into the microbial functional gene repertoire of marine sponges (high microbial abundance, low microbial abundance) and seawater. FEMS Microbiol Ecol 90:832–843. https://doi.org/10.1111/1574-6941.12441.
53. De Mares MC, et al. 2017. Host specificity for bacterial, archaeal, and fungal communities determined for high- and low-microbial abundance sponge species in two genera. Front Microbiol 8:1–13.
54. Sharrar AM, Crits-Christoph A, Méheust R, Diamond S, Starr EP, Cooper JFB. 2020. Bacterial secondary metabolite biosynthetic potential in soil varies with phylum, depth, and vegetation type. mBio 11:e00416-20. https://doi.org/10.1128/mBio.00416-20.
55. Maguire F, Jia B, Gray KL, Lau WYV, Beiko RG, Bringman FSL. 2020. Metagenome-assembled genome binning methods with short reads disproportionately fail for plasmids and genomic islands. Microb Genom 6:mgen000436.
56. Jensen PR. 2016. Natural products and the gene cluster revolution. Trends Microbiol 24:968–977. https://doi.org/10.1016/j.tim.2016.07.006.
57. Tran PN, Yen MR, Chiang CY, Lin HC, Chen PY. 2019. Detecting and prioritizing biosynthetic gene clusters for bioactive compounds in bacteria and fungi. Appl Microbiol Biotechnol 103:3277–3287. https://doi.org/10.1007/s00253-019-09708-z.
58. Ruzzini AC, Clardy J. 2016. Gene flow and molecular innovation in bacteria. Curr Biol 26:R859–R864. https://doi.org/10.1016/j.cub.2016.08.004.
59. Rua CPJ, de Oliveira LS, Froes A, Tschoeke DA, Soares AC, Leomil L, Gregoracci GB, Coutinho R, Hajdu E, Thompson CC, Berlinck RGS, Thompson FL. 2018. Microbial and functional biodiversity patterns in sponges that accumulate bromopyrrole alkaloids suggest horizontal gene transfer of halogenase genes. Microb Ecol 76:825–838. https://doi.org/10.1007/s00248-018-1172-6.
60. Bhat S, Ahrendt T, Dauth C, Bode HB, Shimkets LJ. 2014. Two lipid signals guide fruiting body development of *Myxococcus xanthus*. mBio 5:e00939-13. https://doi.org/10.1128/mBio.00939-13.
61. Gallego-García A, Monera-Girona AJ, Pajares-Martínez E, Bastida-Martínez E, Pérez-Castaño R, Iniesta AA, Fontes M, Padmanabhan S, Elías-Arnanz M. 2019. A bacterial light response reveals an orphan desaturase for human plasmalogen synthesis. Science 366:128–132. https://doi.org/10.1126/science.aay1436.
62. Robinson SL, Terlouw BR, Smith MD, Pidot SJ, Stinear TP, Medema MH, Wackett LP. 2020. Global analysis of adenylate-forming enzymes reveals β-lactone biosynthesis pathway in pathogenic *Nocardia*. J Biol Chem 295:14826–14839. https://doi.org/10.1074/jbc.RA120.013528.
63. Tang W-Z, Yang Z-Z, Sun F, Wang S-P, Yang F, Lin H-W. 2017. Leucanone A and naamine J, glycerol ether lipid and imidazole alkaloid from the marine sponge *Leucandra* sp. J Asian Nat Prod Res 19:691–696. https://doi.org/10.1080/10286020.2016.1240171.
64. Essack M, Bajic VB, Archer JAC. 2011. Recently confirmed apoptosis-inducing lead compounds isolated from marine sponge of potential relevance in cancer treatment. Mar Drugs 9:1580–1606. https://doi.org/10.3390/md9091580.
65. Fedorov SN, Makarieva TN, Guzii AG, Shubina LK, Kwak JY, Stonik VA. 2009. Marine two-headed sphingolipid-like compound rhizochalin inhibits EGF-induced transformation of JB6 P+ Cl41 cells. Lipids 44:777–785. https://doi.org/10.1007/s11745-009-3322-6.
66. Zhang S, Song W, Nothias L-F, Couvillion SP, Webster N, Thomas T. 2022. Comparative metabolomic analysis reveals shared and unique chemical interactions in sponge holobionts. Microbiome 10:22. https://doi.org/10.1186/s40168-021-01220-9.
67. Freeman MF, Gurgui C, Helf MJ, Morinaka BI, Uria AR, Oldham NJ, Sahl H-G, Matsunaga S, Piel J. 2012. Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides. Science 338:387–390. https://doi.org/10.1126/science.1226121.
68. Montalbán-López M, Scott TA, Ramesh S, Rahman IR, van Heel AJ, Viel JH, Bandarian V, Dittmann E, Genilloud O, Goto Y, Grande Burgos MJ, Hill C, Kim S, Koehnke J, Latham JA, Link AJ, Martínez B, Nair SK, Nicolet Y, Rebuffat S, Sahl H-G, Sareen D, Schmidt EW, Schmitt L, Severinov K, Süssmuth RD, Truman AW, Wang H, Weng J-K, van Wezel GP, Zhang Q, Zhong J, Piel J, Mitchell DA, Kuipers OP, van der Donk WA. 2021. New developments in RiPP discovery, enzymology, and engineering. Nat Prod Rep 38:130–239. https://doi.org/10.1039/d0np00027b.
69. Haft DH, Basu MK, Mitchell DA. 2010. Expansion of ribosomally produced natural products: a nitrile hydratase- and Nif11-related precursor family. BMC Biol 8:70–15. https://doi.org/10.1186/1741-7007-8-70.
70. Dunbar KL, Chekan JR, Cox CL, Burkhart BJ, Nair SK, Mitchell DA. 2014. Discovery of a new ATP-binding motif involved in peptidic azoline biosynthesis. Nat Chem Biol 10:823–829. https://doi.org/10.1038/nchembio.1608.

71. Burkhart BJ, Schwalen CJ, Mann G, Naismith JH, Mitchell DA. 2017. YcaO-dependent posttranslational amide activation: biosynthesis, structure, and function. Chem Rev 117:5389–5456. https://doi.org/10.1021/acs.chemrev.6b00623.
72. Haft DH, Basu MK. 2011. Biological systems discovery in silico: radical *S*-adenosylmethionine protein families and their target peptides for posttranslational modification. J Bacteriol 193:2745–2755. https://doi.org/10.1128/JB.00040-11.
73. Li B, Yu JPJ, Brunzelle JS, Moll GN, van der Donk WA, Nair SK. 2006. Structure and mechanism of the lantibiotic cyclase involved in nisin biosynthesis. Science 311:1464–1467. https://doi.org/10.1126/science.1121422.
74. de Oliveira IM, Henriques JAP, Bonatto D. 2007. *In silico* identification of a new group of specific bacterial and fungal nitroreductase-like proteins. Biochem Biophys Res Commun 355:919–925. https://doi.org/10.1016/j.bbrc.2007.02.049.
75. Breil BT, Borneman J, Triplett EW. 1996. A newly discovered gene, *tfuA*, involved in the production of the ribosomally synthesized peptide antibiotic trifolitoxin. J Bacteriol 178:4150–4156. https://doi.org/10.1128/jb.178.14.4150-4156.1996.
76. Cao L, Do T, Link AJ. 2021. Mechanisms of action of ribosomally synthesized and posttranslationally modified peptides (RiPPs). J Ind Microbiol Biotechnol 48:kuab005.
77. Collin F, Maxwell A. 2019. The microbial toxin microcin B17: prospects for the development of new antibacterial agents. J Mol Biol 431:3400–3426. https://doi.org/10.1016/j.jmb.2019.05.050.
78. Travin DY, Watson ZL, Metelev M, Ward FR, Osterman IA, Khven IM, et al. 2019. Structure of ribosome-bound azole-modified peptide phazolicin rationalizes its species-specific mode of bacterial translation inhibition. Nat Commun 10:4563.
79. Kloosterman AM, Shelton KE, Van Wezel GP, Medema MH, Mitchell DA. 2020. RRE-Finder: a genome-mining tool for class-independent RiPP discovery. mSystems 5:e00267-20. https://doi.org/10.1128/mSystems.00267-20.
80. Cao L, Gurevich A, Alexander KL, Naman CB, Leão T, Glukhov E, Luzzatto-Knaan T, Vargas F, Quinn R, Bouslimani A, Nothias LF, Singh NK, Sanders JG, Benitez RAS, Thompson LR, Hamid M-N, Morton JT, Mikheenko A, Shlemov A, Korobeynikov A, Friedberg I, Knight R, Venkateswaran K, Gerwick WH, Gerwick L, Dorrestein PC, Pevzner PA, Mohimani H. 2019. MetaMiner: a scalable peptidogenomics approach for discovery of ribosomal peptide natural products with blind modifications from microbial communities. Cell Syst 9:600–608.e4. https://doi.org/10.1016/j.cels.2019.09.004.
81. Heinilä LMP, Fewer DP, Jokela JK, Wahlsten M, Jortikka A, Sivonen K. 2020. Shared PKS module in biosynthesis of synergistic laxaphycins. Front Microbiol 11:578878–578810. https://doi.org/10.3389/fmicb.2020.578878.
82. Wu X, Flatt PM, Xu H, Mahmud T. 2009. Biosynthetic gene cluster of cetoniacytone A, an unusual aminocyclitol from the endosymbiotic bacterium *Actinomyces* sp. Lu 9419. Chembiochem 10:304–314. https://doi.org/10.1002/cbic.200800527.
83. Fetzner S. 2012. Ring-cleaving dioxygenases with a cupin fold. Appl Environ Microbiol 78:2505–2514. https://doi.org/10.1128/AEM.07651-11.
84. Sekurova O, Pérez-Victoria I, Martín J, Degnes K, Sletta H, Reyes F, Zotchev S. 2016. New deferoxamine glycoconjugates produced upon overexpression of pathway-specific regulatory gene in the marine sponge-derived *Streptomyces albus* PVA94-07. Molecules 21:1131. https://doi.org/10.3390/molecules21091131.
85. Tambadou F, Lanneluc I, Sablé S, Klein GL, Doghri I, Sopéna V, Didelot S, Barthélémy C, Thiéry V, Chevrot R. 2014. Novel nonribosomal peptide synthetase (NRPS) genes sequenced from intertidal mudflat bacteria. FEMS Microbiol Lett 357:123–130.
86. Konstantinou D, Popin RV, Fewer DP, Sivonen K, Gkelis S. 2021. Genome reduction and secondary metabolism of the marine sponge-associated cyanobacterium leptothoe. Mar Drugs 19:298. https://doi.org/10.3390/md19060298.
87. Mevaere J, Goulard C, Schneider O, Sekurova ON, Ma H, Zirah S, Afonso C, Rebuffat S, Zotchev SB, Li Y. 2018. An orthogonal system for heterologous expression of actinobacterial lasso peptides in *Streptomyces* hosts. Sci Rep 8:8232.
88. Chaib De Mares M, Jimenez DJ, Galladino C, Gutleben J, Lebrun LA, et al. 2018. Expressed protein profile of a *Tectomicrobium* and other microbial symbionts in the marine sponge *Aplysina aerophoba* as evidenced by metaproteomics. Sci Rep 8:11795. https://doi.org/10.1038/s41598-018-30134-0.
89. Roume H, Heintz-Buschart A, Muller EEL, Wilmes P. 2013. Sequential isolation of metabolites, RNA, DNA, and proteins from the same unique sample. Methods Enzymol 531:219–236. https://doi.org/10.1016/B978-0-12-407863-5.00011-3.
90. Goethe R, Basler T, Meissner T, Goethe E, Spröer C, Swiderski J, Weiss S, Jarek M, Bunk B. 2020. Complete genome sequence and manual reannotation of *Mycobacterium avium* subsp. *paratuberculosis* strain DSM 44135. Microbiol Resour Announc 9:e00711-20. https://doi.org/10.1128/MRA.00711-20.
91. Bushnell B. 2018. BBMap short read aligner and other bioinformatics tools. https://sourceforge.net/projects/bbmap/.
92. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes *de novo* assembler. Curr Protoc Bioinformatics 70:e102. https://doi.org/10.1002/cpbi.102.
93. Uritskiy GV, DiRuggiero J, Taylor J. 2018. MetaWRAP: a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome 6:158. https://doi.org/10.1186/s40168-018-0541-1.
94. Wu YW, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605–607. https://doi.org/10.1093/bioinformatics/btv638.
95. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 7:e7359-19. https://doi.org/10.7717/peerj.7359.
96. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. 2014. Binning metagenomic contigs by coverage and composition. Nat Methods 11:1144–1146. https://doi.org/10.1038/nmeth.3103.
97. Olm MR, Brown CT, Brooks B, Banfield JF. 2017. DRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J 11:2864–2868. https://doi.org/10.1038/ismej.2017.126.
98. Olm MR, Crits-Christoph A, Diamond S, Lavy A, Matheus Carnevali PB, Banfield JF. 2020. Consistent metagenome-derived metrics verify and delineate bacterial species boundaries. mSystems 5:e00731-19. https://doi.org/10.1128/mSystems.00731-19.
99. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. Bioinformatics 36:1925–1927. https://doi.org/10.1093/bioinformatics/btz848.
100. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2: approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.
101. Letunic I, Bork P. 2019. Interactive Tree of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256–W259. https://doi.org/10.1093/nar/gkz239.
102. Gruber-Vodicka HR, Seah BKB, Pruesse E. 2020. phyloFlash: rapid small-subunit rRNA profiling and targeted assembly from metagenomes. mSystems 5:e00920-20. https://doi.org/10.1128/mSystems.00920-20.
103. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41(Database issue):D590–D596. https://doi.org/10.1093/nar/gks1219.
104. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.
105. Pascal Andreu V, Augustijn HE, van den Berg K, van der Hooft JJJ, Fischbach MA, Medema MH. 2021. BiG-MAP: an automated pipeline to profile metabolic gene cluster abundance and expression in microbiomes. mSystems 6:e00937021. https://doi.org/10.1128/mSystems.00937-21.
106. Rodriguez-R LM, Gunturu S, Tiedje JM, Cole JR, Konstantinidis KT. 2018. Nonpareil 3: fast estimation of metagenomic coverage and sequence diversity. mSystems 3:e00039-18. https://doi.org/10.1128/mSystems.00039-18.
107. Steffen K, Indraningrat AAG, Erngren I, Haglöf J, Becking LE, Smidt H, Yashayaev I, Kenchington E, Pettersson C, Cárdenas P, Sipkema D. 2022. Oceanographic setting influences the prokaryotic community and metabolome in deep-sea sponges. Sci Rep 12:3356. https://doi.org/10.1038/s41598-022-07292-3.