A multiproducer microbiome generates chemical diversity in the marine sponge *Mycale hentscheli*

Michael Rust[a], Eric J. N. Helfrich[a], Michael F. Freeman[a,b,c,1], Pakjira Nanudorn[a,1], Christopher M. Field[a], Christian Rückert[d], Tomas Kündig[a], Michael J. Page[e], Victoria L. Webb[f], Jörn Kalinowski[d], Shinichi Sunagawa[a], and Jörn Piel[a,2]

[a]Institute of Microbiology, ETH Zurich, 8093 Zurich, Switzerland; [b]Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota Twin Cities, St. Paul, MN 55108; [c]BioTechnology Institute, University of Minnesota Twin Cities, St. Paul, MN 55108; [d]Institute for Genome Research and Systems Biology, Center for Biotechnology, Universität Bielefeld, 33594 Bielefeld, Germany; [e]Aquaculture, Coasts and Oceans, National Institute of Water and Atmospheric Research Ltd., 7010 Nelson, New Zealand; and [f]Aquaculture, Coasts and Oceans, National Institute of Water and Atmospheric Research Ltd., 6021 Wellington, New Zealand

Bacterial specialized metabolites are increasingly recognized as important factors in animal–microbiome interactions: for example, by providing the host with chemical defenses. Even in chemically rich animals, such compounds have been found to originate from individual members of more diverse microbiomes. Here, we identified a remarkable case of a moderately complex microbiome in the sponge host *Mycale hentscheli* in which multiple symbionts jointly generate chemical diversity. In addition to bacterial pathways for three distinct polyketide families comprising microtubule-inhibiting peloruside drug candidates, mycalamide-type contact poisons, and the eukaryotic translation-inhibiting pateamines, we identified extensive biosynthetic potential distributed among a broad phylogenetic range of bacteria. Biochemical data on one of the orphan pathways suggest a previously unknown member of the rare polytheonamide-type cytotoxin family as its product. Other than supporting a scenario of cooperative symbiosis based on bacterial metabolites, the data provide a rationale for the chemical variability of *M. hentscheli* and could pave the way toward biotechnological peloruside production. Most bacterial lineages in the compositionally unusual sponge microbiome were not known to synthesize bioactive metabolites, supporting the concept that microbial dark matter harbors diverse producer taxa with as yet unrecognized drug discovery potential.

symbiosis | natural products | biosynthesis | microbiomes | sponges

There is strong evidence that microbiome-derived specialized metabolites play key roles in health, disease, reproductive success, evolutive diversification, and the survival of macroorganisms (1, 2). An important example is defensive symbiosis in which hosts benefit from protective substances synthesized by a microbial partner (3). Since few symbiotic producers have been successfully cultivated, such interactions have to date been uncovered in a relatively small number of cases, but their identification in taxonomically diverse hosts and symbionts suggests that defensive symbiosis is rather prevalent in nature.

As the oldest extant metazoans (4) and prolific sources of bioactive natural products (5), marine sponges offer particularly intriguing opportunities to study symbiotic interactions. Although featuring a simple body plan that lacks specialized tissues, many sponges are complex multispecies organisms containing hundreds to thousands of bacterial phylotypes at remarkable collective cell numbers (6). Little is known about which of the prokaryotes detected by 16S ribosmal RNA (rRNA) gene surveys establishes stable associations with sponges rather than being accumulated by filter feeding or derived from nonspecific colonization (7, 8), and few experimentally validated functions in sponge–bacterial symbiosis have been uncovered (9). For the lithistid sponge *Theonella swinhoei*, a source of an unusually wide array of natural products, we and collaborators recently identified symbiotic "Entotheonella" bacteria of the candidate phylum "Tectomicrobia" as the key producers of bioactive metabolites (10). *T. swinhoei* comprises several sponge variants with diverse and mostly nonoverlapping sets of bioactive metabolites (11). In each of the Japanese chemotypes *T. swinhoei* Y and W, a single symbiont ["*Candidatus* Entotheonella factor" (10, 12, 13) or "*Candidatus* Entotheonella serta" (14), respectively] produces all or almost all of the polyketide and peptide natural products known from the sponges. Each symbiont harbors diverse and almost orthogonal sets of biosynthetic gene clusters (BGCs), providing a rationale for the distinct chemistry of *T. swinhoei* variants. Entotheonella symbionts were also assigned to bioactive metabolites in the lithistid sponges *Discodermia calyx* (15, 16) and a Palauan chemotype of *T. swinhoei*

Significance

Sponges, one of the oldest extant animal phyla, stand out among marine organisms as sources of structurally diverse bioactive natural products. Previous work on chemically rich sponges identified single "superproducer" symbionts in their microbiomes that generate the majority of the bioactive compounds known from their host. Here, we present a contrasting scenario for the New Zealand sponge *Mycale hentscheli* in which a multiproducer consortium is the basis of chemical diversity. Other than the known cocktail of cytotoxins, metagenomic and functional data support further chemical diversity originating from various uncultivated bacterial lineages. The results provide a rationale for distinct patterns of chemical variation observed within sponge species and reinforce uncultured microbes as promising source of compounds with therapeutic potential.

(17). In addition to Entotheonella, which seems to be a widespread producer taxon (10), the cyanobacterium *Oscillatoria spongeliae* has recently been identified as a source of halogenated natural products in dysideid sponges (18–20). Many sponge natural products play suspected or proven roles in chemical defense (6) and have attracted much attention as sources for new therapeutics (5). Commonly, these metabolites are exclusively known from sponges and exhibit pharmacological profiles that impart high drug potential (21). Their low natural abundance, however, represents a major obstacle to drug development that might be overcome by developing bacterial production systems (6).

Similar to *Theonella*, the sponge genus *Mycale* (order Poecilosclerida) is known as a rich and varied source of bioactive substances (22). From New Zealand specimens of *Mycale (Carmia) hentscheli* (23), three groups of cytotoxic polyketides with distinct modes of action have been reported represented by the ribosome-inhibiting (24) contact poison mycalamide A (25) (**1**), the translation initiation inhibitor (26) pateamine A (27) (**2**), and the microtubule inhibitor (28) peloruside A (29) (**3**) (Fig. 1). Peloruside A has attracted attention as a promising anticancer agent since it binds to a microtubule site distinct from inhibitors in clinical use (30). However, attempts to establish a sponge mariculture system for peloruside production were abandoned after invasion of a destructive nudibranch grazer (31). Further impeding drug development, the chemistry of *M. hentscheli* is highly variable, with individual specimens containing all, two, or one of the three polyketide groups (32). Considering the insights gained from the *T. swinhoei* model, understanding microbiome functions in *M. hentscheli* will be crucial for the sustainable use of this resource.

Here, we provide evidence that the known natural products of *M. hentscheli* originate from its microbiota, thus further supporting the hypothesis that chemical functions contributed by bacteria play widespread and fundamental roles in this animal phylum. However, metagenomic data revealed a bacterial consortium in *M. hentscheli* that is, to our knowledge, phylogenetically distinct from that of any previously analyzed sponge. In an opposite scenario to that of *T. swinhoei*, biosynthetic genes are distributed among almost all microbiome members, with a multiproducer consortium being the collective source of the host chemistry. In addition to genes for the three known polyketide types, we identified BGC candidates for various further bioactive compounds previously unknown from this sponge attributed to at least 19 distinct bacteria. The data show that sponges use multiple strategies to acquire and utilize bacterial chemicals. Moreover, they reinforce microbial dark matter as a rich discovery resource that harbors a wide range of previously unrecognized lineages with distinct and biomedically relevant chemistry.

Results

Identification of Gene Candidates for the *M. hentscheli* Polyketides.

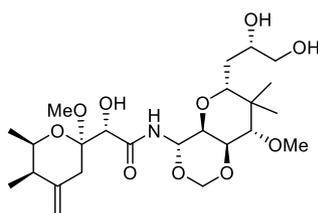We initiated our study on *M. hentscheli* with consideration to the possible enzymatic origin of its known natural products, the mycalamides, pateamines, and pelorusides. Mycalamides belong to the pederin family, a group of defensive polyketides produced in remarkably diverse host–symbiont systems comprising beetles (33), psyllids (34), lichens (35), and sponges (36, 37) as well as free-living bacteria (38, 39) with producers from at least four different phyla. We previously showed that the compounds from non-*Mycale* holobionts are enzymatically generated by a family of polyketide synthases (PKSs) termed *trans*-acyltransferase (*trans*-AT) PKSs (33, 36, 40). For as yet unknown reasons, *trans*-AT PKSs are the predominant enzyme family for complex polyketide biosynthesis in uncultivated symbionts studied to date (41). The structure of pateamine contains three methyl groups (at C6, C13, and C20) (Fig. 1) at positions that suggest attachment by a β-branching mechanism (42). This feature is ubiquitous for *trans*-AT PKSs but rare in *cis*-acyltransferase (*cis*-AT) PKS pathways, the second large enzymatic source of complex polyketides. Peloruside lacks structural moieties that indicate its PKS type. Hypothesizing that *trans*-AT PKSs generate at least two of the three polyketide series, we performed metagenomic sequencing.

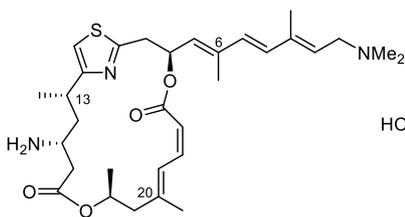### A Medium-Sized Microbiome Containing Diverse Natural Product Biosynthesis Genes.

DNA was extracted from different *M. hentscheli* specimens that were positive for at least two of the three polyketides. Fast degradation of some DNA samples during purification posed a major challenge for downstream processing. Ultimately, we obtained high-molecular weight DNA from two sponge specimens designated as Myc1 (sponge ID: 1MJP40-24) and Myc2 (sponge ID: 1MJP3-79.12) containing all three polyketides (*SI Appendix*, Table S1), which was subjected to Illumina sequencing. Reads were assembled separately for the two sponge samples using metaSPAdes (43) (*SI Appendix*, Table S2). We obtained initial insights into the general biosynthetic potential of the *M. hentscheli* microbiome by analyzing contigs >1,500 bp with the automated BGC detection tool antiSMASH (44) in combination with extensive manual analyses to identify noncanonical BGCs not detected by the software. The data revealed a plethora of BGC-containing contigs from diverse biosynthetic families (*SI Appendix*, Tables S3 and S4). Since there is a close correlation between the module architecture of a PKS and the polyketide structure (45), assumptions about the core BGC are possible if the structure is known and vice versa. Gratifyingly, BGC regions on some contigs correlated well with the structures of mycalamide and pateamine (*Insights into Mycalamide, Pateamine, and Peloruside Biosynthesis*). In addition, many other contigs contained PKS or NRPS genes that could not be assigned to known compounds, suggesting that the *M. hentscheli* microbiome harbors a greater biosynthetic potential than previously expected. Since repetitive modular PKS regions prevented the assembly of some BGC regions, we also attempted sequencing one of the sponge specimens with a nanopore-based sequencing platform (MinION). However, this step did not improve the overall contig length and quality of the Illumina dataset.

Fig. 1. Representative natural products from the marine sponge *M. hentscheli*. Structures of selected congeners for the three polyketide families are shown. The methyl groups at numbered positions of pateamine A are thought to be introduced by a β-branching mechanism common for *trans*-AT PKSs.

Mycale hentscheli   Mycalamide A (**1**)   Pateamine A (**2**)   Peloruside A (**3**)

Assembled contigs were binned to gain insights into the diversity of microorganisms and their BGC content. The relatively limited number of bins (14 for Myc1 and 20 for Myc2 with >60% estimated genome completeness) (*SI Appendix*, Table S5) revealed a microbiome of much lower diversity in *M. hentscheli* than previous 16S rRNA studies had suggested for other *Mycale* species (46) or for example, *T. swinhoei* (47). Of 14 genomes in the Myc1 dataset, 12 were at least partially present in the Myc2 dataset. The 22 binned genomes were taxonomically classified (https://github.com/Ecogenomics/GTDBTk) and affiliated with 11 bacterial phyla (Fig. 2); 14 of 22 closest bacterial neighbors identified by whole-genome phylogeny originated from cultivation-independent sequencing studies (*SI Appendix*, Fig. S1). Based on the chemical richness of *M. hentscheli*, we had initially hypothesized that it might contain an Entotheonella relative as a talented producer. However, no Entotheonella-related genes were detected in our data, and none of the *Mycale* bins were assigned to the candidate phylum Tectomicrobia to which Entotheonella belongs (10). Further analysis showed that, in contrast to the Entotheonella superproducers in *T. swinhoei*, BGCs are distributed across a large phylogenetic range of bacteria in *M. hentscheli*, with most microbiome members harboring only few biosynthetic pathways (Fig. 2).

The bins with a >90% estimated genome completeness were analyzed for the presence of central metabolism and amino acid biosynthesis pathways (*SI Appendix*, Table S6). Although some of these pathways are partially or completely absent in some bins, we did not observe any cases of extreme genome reduction as found for various intracellular symbionts (48, 49). Furthermore, we analyzed these bins for the presence of clusters of orthologous genes associated with symbiosis (*SI Appendix*, Table S7). All bins harbor several of these "symbiosis factors" with eukaryotic-like proteins, such as ankyrin repeats (Clusters of Orthologous Groups identifier COG0666), tetratricopeptide repeats (COG0457, COG0790), and WD40 proteins (COG2319, COG1520), being the most abundant.

Insights into Mycalamide, Pateamine, and Peloruside Biosynthesis. The structural similarity of pederin-type compounds (diaphorin, nosperin, onnamide A, pederin, psymberin) is reflected in their shared PKS architectures (41). This correlation in combination with the binning data facilitated the assignment of partial BGCs in the metagenomic dataset to the mycalamide pathway. In an initial assembly, we identified four contigs in the same bin with protein architectures that match the predicted mycalamide biosynthesis. These contigs were subsequently connected by PCR to yield the complete *myc* BGC (Fig. 3*A* and *SI Appendix*, Table S8).
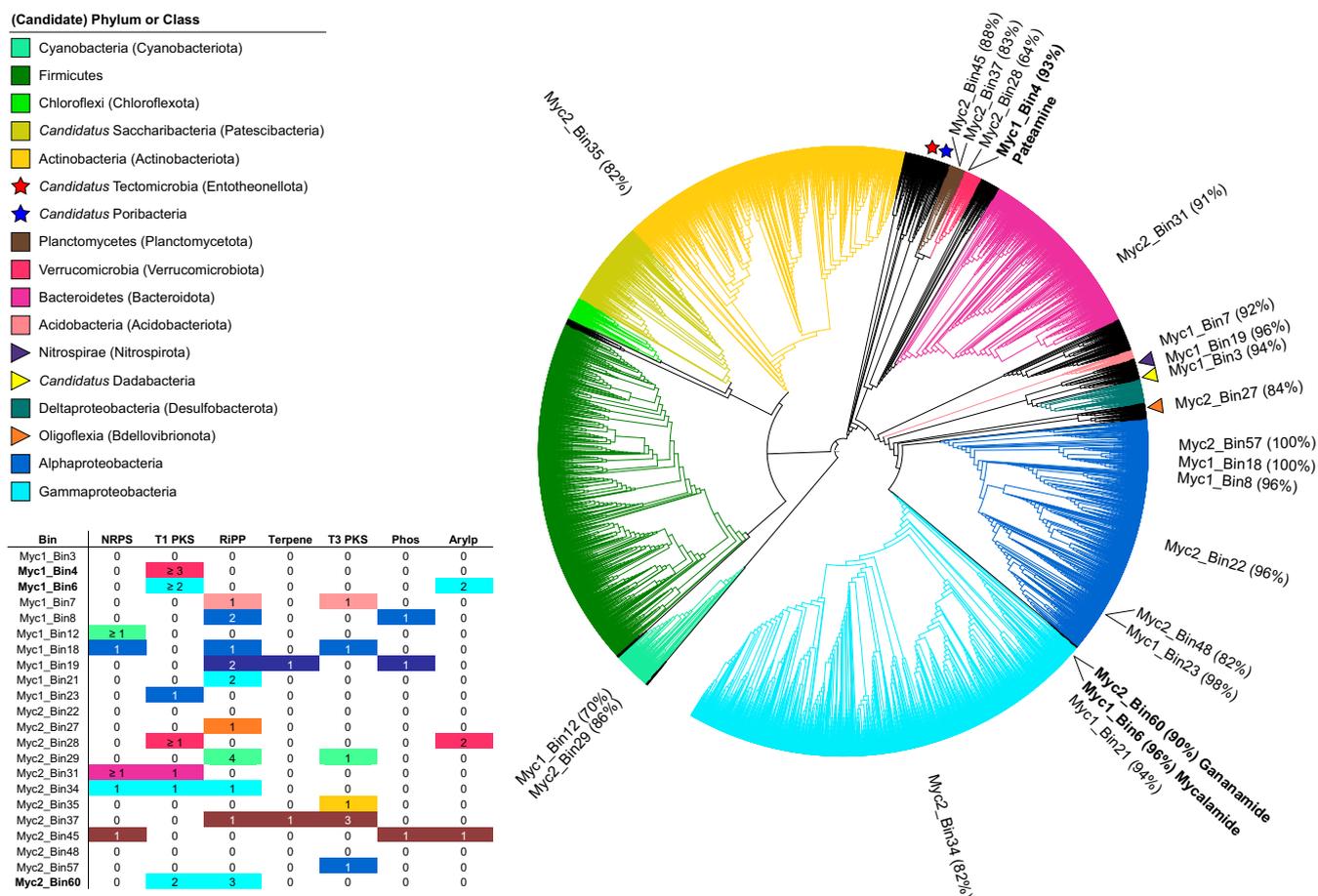
Fig. 2. Taxonomic classification of 22 *Mycale* bins and distribution of identified BGCs. Maximum likelihood placement of the 22 bins with >60% estimated genome completeness (values given in parentheses behind each bin number) in the GTDB-Tk reference tree (58) consisting of 17,435 leaves. Cyanobacteria were used as outgroup. Colored stars denote the placement of the two candidate phyla "Poribacteria" and Tectomicrobia known from previous sponge microbiome studies. Colored triangles denote positions of *Mycale* bins affiliated with phyla with few known genomes (Nitrospirae, *Candidatus* Dadabacteria, Oligoflexia). The three bins harboring the mycalamide, pateamine, and gananamide BGCs are shown in bold. The number of BGCs per bin is listed in the table and color coded according to the bin's assigned phylum. The symbol ≥ is used if multiple biosynthetic contigs were found that might belong to the same pathway. Contigs encoding NRPS–PKS hybrid systems were listed as PKSs. Arylp, arylpolyene; Phos, phosphonate. NRPS, nonribosomal peptide synthetase; RiPP, ribosomally synthesized and posttranslationally modified peptide; T1, type 1; T2, type 2.

Bin	NRPS	T1 PKS	RiPP	Terpene	T3 PKS	Phos	Arylp
Myc1_Bin3	0	0	0	0	0	0	0
Myc1_Bin4	0	≥ 3	0	0	0	0	0
Myc1_Bin6	0	≥ 2	0	0	0	0	2
Myc1_Bin7	0	0	1	0	1	0	0
Myc1_Bin8	0	0	2	0	0	1	0
Myc1_Bin12	≥ 1	0	0	0	0	0	0
Myc1_Bin18	1	0	1	0	1	0	0
Myc1_Bin19	0	0	2	1	0	1	0
Myc1_Bin21	0	0	2	0	0	0	0
Myc1_Bin23	0	1	0	0	0	0	0
Myc2_Bin22	0	0	0	0	0	0	0
Myc2_Bin27	0	0	1	0	0	0	0
Myc2_Bin28	0	≥ 1	0	0	0	0	2
Myc2_Bin29	0	0	4	0	1	0	0
Myc2_Bin31	≥ 1	1	0	0	0	0	0
Myc2_Bin34	1	1	1	0	0	0	0
Myc2_Bin35	0	0	0	0	1	0	0
Myc2_Bin37	0	0	1	1	3	0	0
Myc2_Bin45	1	0	0	0	0	1	1
Myc2_Bin48	0	0	0	0	0	0	0
Myc2_Bin57	0	0	0	0	1	0	0
Myc2_Bin60	0	2	3	0	0	0	0

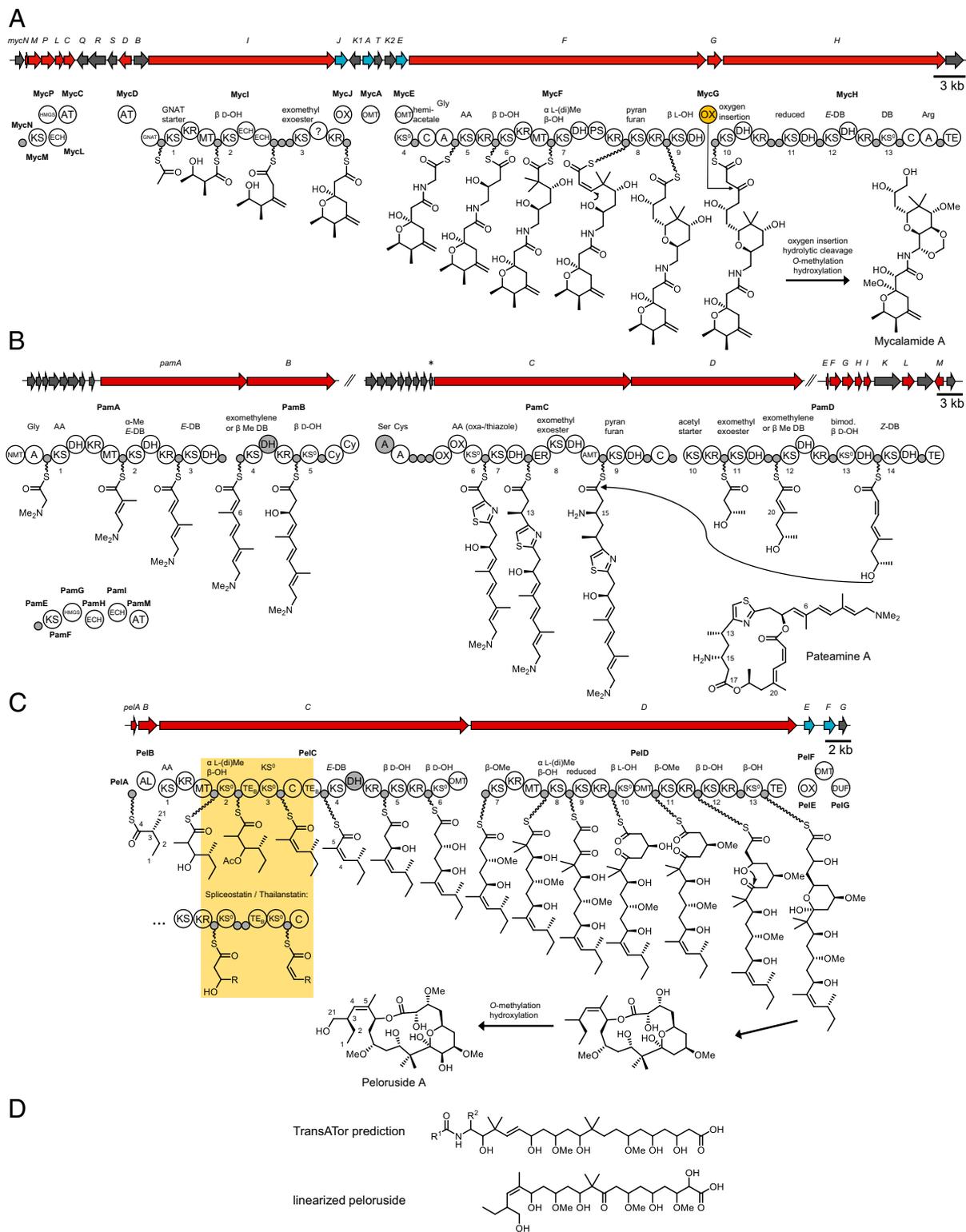**Fig. 3.** Gene clusters and biosynthetic models for mycalamide, pateamine, and peloruside. Core PKS/NRPS genes are shown in red, tailoring biosynthetic genes in blue, and genes with unknown function in gray. Gaps between domains denote protein boundaries. Biosynthetic intermediates are shown tethered to the ACP/peptidyl carrier protein domains (small gray circles). Predicted substrates are shown above the KS and A domains (exomethyl/exoester refers to a clade of KSs that mainly accept intermediates containing β-branched methyl groups or a β-branched ester in case of bryostatin). Domains depicted in gray are predicted to be nonfunctional. (*A*) The *myc* gene cluster and biosynthetic model for mycalamide. The oxygenase thought to introduce oxygen into the growing polyketide chain is highlighted in orange. (*B*) The *pam* gene cluster and biosynthetic model for pateamine. Double slashes denote separate contigs. (*C*) The *pel* gene cluster and biosynthetic model for peloruside. (*D*) TransATor-based structure prediction for the putative *pel* BGC. The linearized peloruside structure is shown for comparison. AA, amino acid; AL, acyl-CoA/ACP ligase; DB, double bond; DUF, domain of unknown function 955; ECH, enoyl-CoA hydratase; GNAT, GCN5-related *N*-acetyl transferase superfamily; HMGS, 3-hydroxy-3-methylglutaryl-CoA synthase homolog; KR, ketoreductase; KS⁰, non-elongating KS; MT, *C*-methyltransferase; OMT, *O*-methyltransferase; OX, oxidoreductase; PS, pyran synthase. *Transposase gene.

Additionally, the BGC architecture was confirmed during an improved assembly in which the complete locus was present on a single contig. The architecture of the hybrid *trans*-AT PKS–NRPS assembly line closely resembles those of the pederin (33) and onnamide (36) PKSs. An interesting feature shared with the pederin system is the large PKS gene *mycH* that does not correspond to any part of the polyketide structure. MycG encoded directly upstream of *mycH* belongs to a group of PKS-associated monooxygenases, for which biochemical data support a function as oxygen-inserting Baeyer–Villigerase acting on growing polyketide chains (50). In the case of pederin and mycalamide, hydrolytic cleavage of the resulting ester moiety would be in agreement with the oxygenated polyketide terminus and missing polyketide portion. Based on the 16S rRNA gene sequence in the *myc* BGC-containing bin, the producer is affiliated with the marine group UBA10353 [94% sequence identity to its nearest neighbor (51)], an uncultured gammaproteobacterial taxon that was previously not known as a source of pederin-type compounds or other natural products. In agreement, the nearest neighbor in the genome-based phylogeny is an uncultivated gammaproteobacterium from the UBA10353 order associated with a glass sponge (*SI Appendix*, Fig. S1) (52). The name "*Candidatus* Entomycale ignis" is proposed for the mycalamide producer.

Pateamine contains a rare *N*-methylated glycine starter. As a match for this diagnostically useful moiety, we identified in another bin the NRPS–PKS gene *pamA* encoding a predicted glycine-specific NRPS module with an *N*-methyltransferase domain. To further interrogate a role in pateamine biosynthesis, the downstream PKS modules were analyzed by TransATor, a recently developed web application that allows functional assignment of *trans*-AT PKS pathways (53). The software uses the correlation between phylogenetically similar ketosynthase (KS) domains and incoming intermediates carrying similar chemical moieties in the α- to γ-region around the thioester (45). Predicted intermediates for the PKS modules in PamA showed good agreement with the C1 to C10 moiety of pateamine (*SI Appendix*, Table S9). Two additional NRPS Cy domains at the C terminus of PamB suggested the presence of an oxazol(in)e or thiazol(in)e moiety (54), consistent with the pateamine thiazole unit. NRPS cyclization modules for aromatic azoles normally also contain serine- or cysteine-specific adenylation (A) domains and an oxidoreductase domain. A single incomplete PKS gene, *pamC*, encoding such domains was identified in the same bin on another contig. We were unable to connect the two contigs by PCR or reassembly, and since *pamC* was preceded by a series of genes unrelated to natural product biosynthesis, the contigs might be located in different genome regions. Candidates for further *pam* PKS genes were identified on three additional contigs in that bin, which encoded enzymes consistent with the predicted missing biosynthetic steps. Successful connection of these contigs by PCR generated a continuous 50-kb fragment ending with a thioesterase (TE) region (Fig. 3B and *SI Appendix*, Table S10). PamC contains various rare features; however, these are consistent with pateamine biosynthesis and include 1) an enoylreductase domain associated with the β-branching module for KS7 and needed to reduce an initially generated sp^2 center at C13, 2) an aminotransferase domain in the downstream module that matches the amino group at C15, and 3) as candidates for the C17 ester function that interrupts the pateamine chain, an NRPS condensation (C) domain in the terminal PamC module and an N-terminal KS on PamD predicted to accept acetyl starters. A similar C domain is also proposed to introduce an ester moiety in the malleilactone (= burkholderic acid) pathway (55, 56). These features suggest that pateamine is biosynthesized by esterification of two separate polyketide chains rather than a mycalamide-type Baeyer–Villiger oxidation. Pateamine contains three β-branches at C6, C13, and C20, for which accessory biosynthetic genes, termed β-branching cassette, are required. These were identified on a third contig that could not

be connected to the other fragments by PCR (*pamEFGHIM*). The 16S rRNA gene of the pateamine producer, with the proposed name "*Candidatus* Patea custodiens" (from the Latin word custos for guardian), was verified by PCR reamplification and sequencing. SILVA (51) analysis of the 16S rRNA gene revealed low sequence identity (90.3%) to a member of the Kiritimatiellaeota, a recently proposed phylum previously assigned to Verrucomicrobia (57). This affiliation is consistent with whole genome-based phylogeny that identified members of the Kiritimatiellaeota as closest neighbors (*SI Appendix*, Fig. S1) (58). Like for the mycalamide-assigned UBA10353 taxon, Kiritimatiellaeota were previously not known as a natural product source.

In the unbinned metagenomic fraction, a 55-kb contig encoding a seemingly complete *trans*-AT PKS assembly line (*pel* PKS) (Fig. 3C and *SI Appendix*, Table S11) attracted our attention. For the large PKS portion between KS4 and the terminal TE, the TransATor core structure prediction suggested a polyketide structure that is almost identical to the peloruside portion covering the macrocycle (Fig. 3D and *SI Appendix*, Table S12). In contrast, the first PKS modules contained numerous noncanonical features that made a functional assignment challenging. More detailed analysis revealed that the unusual domain series starting with KS2 (KS^0-acyl carrier protein [ACP]-TE$_B$-KS^0-ACP-C) occurs in an almost identical sequence (KS^0-ACP-ACP-TE$_B$-KS^0-ACP-C) in the spliceostatin and thailanstatin PKSs (59, 60), where it was assigned to a Z double bond, a feature also found at the matching peloruside moiety (C4 to C5). Interestingly, all three PKSs contain members of a phylogenetically distinct group of internal TE-like domains, previously misannotated as a dehydratase (DH) in the spliceostatin PKS, for which we proposed the name TE$_B$. Work to be published elsewhere on a TE$_B$ homolog from the oocydin PKS suggests that the domain attaches an acetyl side chain to a hydroxyl function. In the biosynthesis of peloruside and the statins, this modification might facilitate elimination as a DH-independent mechanism to introduce double bonds. Moving further upstream along the *pel* PKS, the module containing KS1 is predicted to introduce an α-methyl-β-hydroxyl moiety based on its domain architecture and the phylogeny of KS2, which again agrees with the peloruside moiety. This assignment suggests two alternatives for the origin of the remaining C$_5$ unit composed of C1 to C4 and C21 of peloruside: incorporation of a 2-methylbutyryl starter or iterative action of the KS1 module to elongate an acetyl starter twice.

To obtain initial biochemical insights into the *pel* pathway, we focused on the monodomain protein PelB. Its resemblance to acyl-CoA (coenzyme A)/ACP ligases initially suggested that it might load the starter unit onto the free-standing ACP PelA. We tested its substrate preference by producing PelB in *Escherichia coli* and performing in vitro assays with a range of carboxylic acids (4 to 17) and CoA (*SI Appendix*, Fig. S2A). Monitoring reactions by high-performance liquid chromatography–mass spectrometry (HPLC-MS) revealed that PelB efficiently converted acetic acid (4) to acetyl-CoA but did not accept acids 5 to 17 (*SI Appendix*, Fig. S2B). Unexpectedly, however, coincubations with PelA failed to generate acylated ACP species for acetate or any other test substrate. One explanation for this result might be that PelB provides its acetyl substrate not to the ACP but to another acceptor, such as the TE$_B$. PelA could be loaded with an isobutyryl unit by an as yet unknown enzyme, a scenario that is also supported by the predicted specificity of KS1 for an amino acid-type substrate rather than acetyl. Experiments to clarify this issue are underway but out of the scope of the current study considering the challenge to characterize an assembly line with multiple noncanonical features from an uncultivated bacterium. Since the peloruside sequence was not binned, it might be located on a plasmid with a different tetranucleotide frequency than the rest of the producer genome.
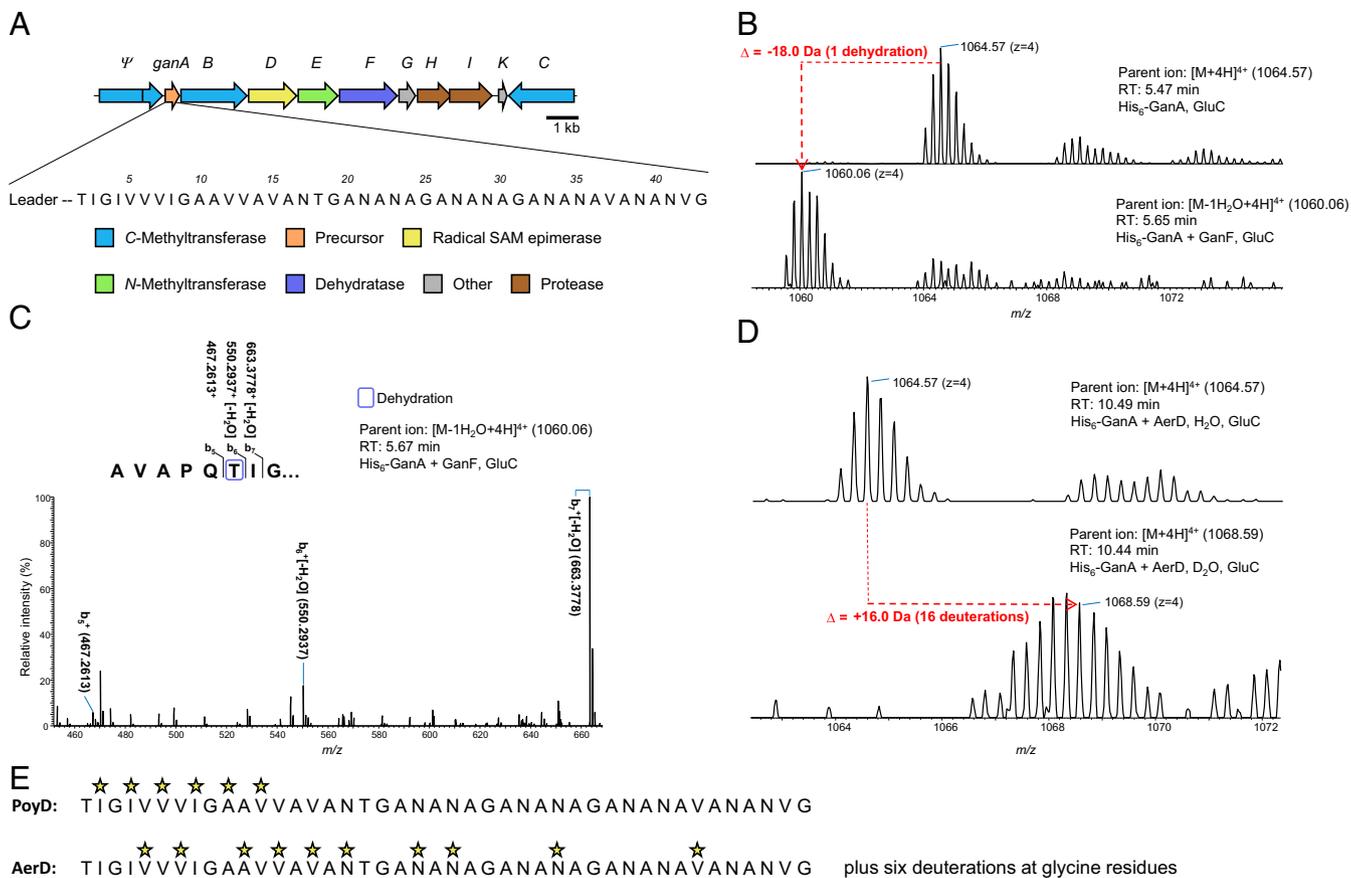
High Natural Product Potential Collectively Encoded in Diverse Microbiome Members. Deeper analyses of the metagenome revealed many additional BGCs for PKSs, NRPSs, ribosomally synthesized and posttranslationally modified peptides (RiPPs), and other compounds that could not be assigned to reported *M. hentscheli* metabolites (Fig. 4). In the bin of "Patea custodiens," the Kiritimatiellaeota pateamine producer, we identified a small *trans*-AT PKS cluster (Fig. 4, cluster 3) that appears complete based on the presence of loading and termination modules. The terminal PKS modules are architecturally almost identical to those of the psymberin PKS (37), suggesting an isocoumarin-type compound as the product. The same bin harbors four additional *trans*-AT PKS contigs (Fig. 4, clusters 1, 2, 4, and 5), including a large 34-kb fragment ending with a TE. "P. custodiens," therefore, likely generates at least two additional polyketides that are currently unknown. The bin of the mycalamide producer "Entomycale ignis" harbors two small *cis*-AT PKS contigs (Fig. 4, clusters 7 and 8), suggesting the production of an additional polyketide. Another bin (assigned to "Caria hoplita") (*Functional Insights into Polytheonamide-Type Pathway in the Nitrosococcaceae Symbiont Caria hoplita*) contains a small, seemingly complete NRPS/PKS hybrid cluster (Fig. 4, cluster 9). In addition, the analyses revealed multiple *trans*-AT PKS contigs in the unbinned fraction (Fig. 4, clusters 14 to 18), including one

encoding a large *trans*-AT PKS (Fig. 4, cluster 14) and a glycosyltransferase, suggesting a glycosylated polyketide. Another small and apparently complete *trans*-AT PKS BGC (Fig. 4, cluster 16) encodes an architecturally unusual PKS containing a domain of unknown function and pyridoxal phosphate-dependent enzyme that are similar to those assigned to sulfur heterocycle biosynthesis in leinamycin (61) as well as a PedG/MycG-type Baeyer–Villigerase homolog. In summary, the *M. hentscheli* microbiome is particularly rich in *trans*-AT PKSs, pathways that are often found in uncultivated symbiotic bacteria, but also harbors various BGCs from additional natural product classes (*SI Appendix*, Tables S3 and S4).

Functional Insights into Polytheonamide-Type Pathway in the Nitrosococcaceae Symbiont Caria hoplita. An unexpected orphan locus identified in the *M. hentscheli* metagenome (Fig. 5*A* and *SI Appendix*, Table S13) closely resembles the polytheonamide (*poy*) cluster of "E. factor," the multiproducer of the sponge *T. swinhoei* Y (62). Polytheonamides (63) are extraordinarily complex and rare cytotoxic peptides generated by a RiPP pathway involving 49 posttranslational modifications (62, 64). The unmodified precursor consists of an N-terminal leader region and a C-terminal core that is processed by maturases during biosynthesis and ultimately cleaved off by proteolysis. For five of six polytheonamide maturation

Fig. 4. Orphan PKS gene clusters. Core PKS/NRPS genes are shown in red, tailoring biosynthetic genes in blue, and genes with unknown function in gray. Gaps between domains denote protein boundaries. Small gray circles represent ACP/peptidyl carrier protein domains. Clusters 1 to 6 were identified in the pateamine producer Patea custodiens, clusters 7 and 8 were identified the mycalamide producer Entomycale ignis, cluster 9 was identified in the gananamide producer Caria hoplita, and clusters 14 to 18 were identified in the unbinned fraction. DUF, domain of unknown function 2156; FAAL, fatty acyl-AMP ligase; FkbH, hydroxylase; FkbM, methyltransferase; FT, formyltransferase; GT, glycosyltransferase; KR, ketoreductase; KS⁰, nonelongating KS; OX, oxidoreductase; PLP, pyridoxal phosphate-dependent enzyme; TD, terminal domain.

Fig. 5. The *gan* BGC and functional characterization of modifying enzymes. (*A*) The *gan* cluster with genes color coded according to known enzymatic functions of the polytheonamide homologs. The core sequence of the precursor GanA (43 amino acids) containing the repetitive GANANA motif is shown; Ψ indicates rSAM *C*-methyltransferase pseudogene. (*B*) MS spectra of the GluC-digested precursor GanA produced in *E. coli* without (*Upper*) and with (*Lower*) the dehydratase GanF. (*C*) Tandem mass spectrometry (MS²) spectrum of the GluC-digested precursor GanA coproduced with GanF, localizing the dehydration to Thr1 at the N terminus of the core. (*D*) MS spectra of GluC-digested precursor GanA coproduced with the epimerase AerD in H_2O and D_2O. A mass shift corresponding to a total of 16 deuterations was detected. (*E*) Localization of epimerized amino acids (star symbols) in the core sequence of GanA introduced by the epimerases PoyD and AerD. RT, retention time.

enzymes, homologs were identified in the *M. hentscheli* BGC (*gan* cluster). These comprise a radical *S*-adenosyl methionine (rSAM) epimerase (generating 18 D-amino acids in polytheonamides), a Ser/Thr DH (*poy*: Thr1 dehydration), 2 rSAM *C*-methyltransferases (*poy*: 17 *C*-methylations), and an *N*-methyltransferase (*poy*: 8 Asn side-chain *N*-methylations). The core sequence of the precursor GanA likewise resembles the *poy* core in length and composition but contains a characteristic GANANA repeat. We, therefore, provisionally named the orphan natural product gananamide.

To test whether the *gan* cluster is functional, we coexpressed in *E. coli* genes for the DH homolog GanF and the precursor GanA, carrying an added N-terminal His₆ tag. As in the polytheonamide pathway, GanF efficiently dehydrated the threonine residue at core position 1 (Fig. 5 *B* and *C*). Further coexpression trials with the epimerase gene *ganD*, however, failed to generate active enzyme. Since we previously showed that the core sequence largely dictates the D-amino acid pattern introduced by rSAM epimerases (65), we performed individual coexpression experiments with PoyD and AerD to obtain clues about D-residues in gananamide. AerD is a recently identified epimerase with improved processitvity in *E. coli* from the polytheonamide-type aeronamide pathway in *Microvirgula aerodenitrificans* (66). For both epimerase homologs, additional peaks with different retention times appeared in the HPLC-MS data (*SI Appendix*, Fig. S3). Subsequent application of a previously developed method that permits quantification and localization of D-residues through

deuteration (ODIS [orthogonal D_2O-based induction system]) (67) revealed 6 epimerized amino acids for PoyD at the N-terminal core portion of GanA and 10 epimerizations for AerD (Fig. 5 *D* and *E* and *SI Appendix*, Figs. S4–S6). Experiments to identify the mature natural product in *M. hentscheli* are underway.

The partial 16S rRNA gene of the proposed gananamide producer, for which we suggest the name "*Candidatus* Caria hoplita" (based on hoplites, armed foot soldiers that played a role in defending the ancient region Caria during the battle of Mycale), was completed by PCR amplification and sequencing. It was affiliated with the Nitrosococcaceae and has the highest sequence identity of 94.7% to an uncultured sediment bacterium belonging to this family. The closest neighbor based on available whole genomes is an unclassified gammaproteobacterium from a water purification plant metagenome (68).

Discussion

With more than 8,000 species being globally distributed, sponges make up a successful and evolutionarily ancient animal phylum (69). Many sponges harbor remarkably diverse microbial communities, for which various contributions to the host physiology and ecology have been proposed or in fewer cases, experimentally demonstrated (9). Among the latter is the provision of bioactive natural products that can be present as rich arrays in some holobionts (70). Such studies have identified single members within more diverse microbiomes that generate some or all

of the specialized metabolites known from the host (10, 14, 20, 48). Here, we reveal a contrasting scenario in which the source of three distinct cytotoxin families, the mycalamides (25), pateamines (27), and pelorusides (29), is a complex chemistry-based symbiosis in the poecilosclerid sponge *M. hentscheli*. Rather than production being localized in individual bacteria within BGC-depleted microbiomes (10, 14, 48), the data support a producer consortium in which multiple phylogenetically diverse members contribute to the overall rich chemistry of the holobiont. Other than BGCs assigned to the three known polyketide classes, this microbiome was found to contain numerous additional loci from distinct natural product families. Most of these BGCs appear to be intact, and functional studies on the cryptic polytheonamide-type gananamides provide additional evidence that richer chemistry is to be expected. Current work in our group and collaborating groups aims at revealing the identity of these unknown metabolites.

M. hentscheli has attracted much attention based on the therapeutic potential of its compounds, with pelorusides being particularly promising anticancer drug candidates (71). One of several obstacles in the pharmaceutical development (31) of this resource has been the high chemical variability among sponge specimens regarding the presence of individual compounds (32). This complex variation contrasts with the association with distinct "superproducers" as in the case of *T. swinhoei* with a complete switch of metabolic profiles (10, 14) or variations of *Prochloron* phylotypes associated with tropical ascidians (72, 73). In *M. hentscheli*, a multiproducer consortium of variable composition is the likely reason for the diverse chemotypes. With knowledge on producers and their BGCs available, further development of *M. hentscheli* compounds could be expedited by permitting rapid PCR profiling of specimens, targeted isolation of producers from the host or alternative sources (66, 74), or the establishment of heterologous expression systems.

Among all chemically assigned BGCs, the *pel* cluster assigned to pelorusides is the only one that could not be linked to a specific producer by binning, perhaps because it is located on a plasmid. To address this issue, collection of fresh specimens will be required to localize *pel* genes by in situ hybridization or single-cell genomics. Biosynthetic assignments of the architecturally aberrant *pel* BGC as well as the fragmented pateamine (*pam*) PKS loci were possible by retrobiosynthetic dissection and analysis of the KS domains (45, 53), showcasing the value of in silico biosynthetic predictions using *trans*-AT PKS correlations. These correlations provide reasonably sound biosynthetic models that can be used as basis for further functional analyses and heterologous expression studies. The functions of the acetyl transferase PelB within the peloruside assembly line and the unusual domains at the start of PelC are currently under investigation in our laboratory. The relatively small size and apparent plasmid-based localization of the six-gene *pel* cluster and the pharmacological relevance of peloruside render this pathway a promising system for heterologous expression studies.

Metagenomic and single-cell datasets provide a valuable data resource to study how natural products evolve and disperse in symbiotic systems, and these are beginning to reveal intriguing patterns (75). The identification of the mycalamide producer Entomycale ignis adds a further bacterial lineage to an already astonishing diversity of organisms producing pederin-type compounds, reinforcing intriguing questions about the evolution and dispersal of these natural products. Previously identified producers belong to alpha- (38), beta- (34) and gammaproteobacteria (33) as well as two Cyanobacteria (35, 39) and Tectomicrobia (10), and they are mostly symbionts of unrelated hosts comprising sponges (10), beetles (33), psyllids (34), and a lichen fungus (35). The data support extensive horizontal gene transfer and retention in extraordinarily diverse host–symbiont systems, including intracellular organelle-like bacteria with minimalistic genomes (34) as well as multicellular prokaryotes featuring genomes of around

10 Mb (10, 14). Pederin-type metabolites might thus represent ancient, "symbiotically privileged" natural products that have driven the evolution of multiple symbioses through host protection (76, 77). However, more recently, pederin-type compounds were also reported from two free-living bacteria (38, 39).

"Microbial dark matter," which comprises numerous deep-branching clades lacking cultivated members, has been proposed as a rich and largely untapped resource of novel chemistry (78, 79). Studies on microbiomes of marine invertebrates provide direct experimental support for this hypothesis: in addition to the talented producer taxon Entotheonella within an uncultured candidate phylum (10, 14), they have uncovered lineages, such as "Endobugula" (80), "Endohaliclona" (48), "Endolissoclinum" (49), "Didemnitutus" (81), and "Endobryopsis" (82), as new natural product sources from uncultivated life. The current study further expands the range of producers by members of Kiritimatiellaeota (57), the UBA10353 taxon, Nitrosococcaceae, Verrucomicrobia, and other groups. This growing prokaryotic diversity collectively generates a wide range of bioactive compounds with chemical scaffolds that were largely unknown from conventionally screened bacterial taxa, such as actinomycetes. The data suggest that, with cultivation-independent studies becoming more routine, continued functional exploration of microbial dark matter will substantially change our understanding of bacterial specialized metabolism.

Materials and Methods

Instrumentation. Ultraperformance liquid chromatography–heated electrospray ionization mass spectrometry was performed on a Thermo Scientific Q Exactive mass spectrometer coupled to a Dionex Ultimate 3000 UPLC system.

Sample Collection. *M. hentscheli* specimens were collected in Pelorus Sound at the northern end of the South Island of New Zealand and stored in *RNA*later as described previously (83).

Isolation of Metagenomic DNA. Metagenomic DNA was isolated from 1 g of sponge sample as described previously (84) with slight modifications. Briefly, the sponge material was ground to a fine powder under liquid nitrogen, transferred to a 50-mL falcon tube containing 10 mL sponge lysis buffer, and incubated at 60 °C for 20 min. The sample was extracted two times with phenol/chloroform/isoamyl alcohol (25:24:1, vol/vol/vol), and the aqueous phase was extracted with chloroform. The DNA was precipitated by adding 1/10th vol of 3 M sodium acetate (pH 7) and 1 vol isopropanol. The sample was gently mixed by inverting the tube several times and then, incubated at room temperature for 30 min. The tube was centrifuged at 10,000 × *g* and 4 °C for 30 min. The supernatant was removed, and the DNA pellet was washed twice with 70% ice-cold ethanol. The sample was centrifuged at 10,000 × *g* and 4 °C for 20 min after each washing step. The pellet was air dried for 5 min and then, resuspended in 500 μL prewarmed Tris·HCl buffer (5 mM, pH 8.5). The reasons for the fast degradation of some DNA samples are currently unknown. A fast downstream processing after the sponge lysis step was crucial for obtaining high-molecular weight DNA.

Metagenome Sequencing, Assembly, and Binning. Sequencing libraries (2× 250-bp paired-end reads) were prepared from purified metagenomic DNA and sequenced using an Illumina HiSeq2500 platform. BBDuk (v37.55; Joint Genome Institute) was first used in right-trimming mode with a kmer length of 23 down to 11 and a Hamming distance of one to filter out sequencing adapters. A second pass with a kmer length of 31 and a Hamming distance of one was used to filter out PhiX sequences. A third and final pass performed quality trimming on both read ends with a Phred score cutoff of 15 and an average quality score cutoff of 20, with reads under 30 bp or containing Ns subsequently rejected.

For the initial assembly (fast and memory efficient), all paired-end reads from both sponge specimens (Myc1 and Myc2) were combined and assembled using MEGAHIT (85) (v1.1.1) with "–k-list 39,59,79,99" and otherwise, default parameters. Assembly with metaSPAdes (43) (usually better assembly statistics) required more random access memory (RAM) than the 3 terabytes that we had available, and therefore, the paired-end reads of each read set were first normalized to reduce the complexity of the assembly graph. BBNorm (v37.55; Joint Genome Institute) was used with a default kmer length of 31, a minimum depth of two, and a target depth of 80 to down sample sequences of high depth and filter out unique kmers.

The normalized paired-end and unnormalized singleton reads of each read set were assembled using metaSPAdes (v3.11.0) without the error correction module but otherwise, default parameters. Scaffolds smaller than 1,500 bp were then filtered out. The quality-controlled paired-end reads were aligned to the assembled metaSPAdes scaffolds using BWA (86) (v0.7.15-r1140), and the alignments were sorted by SAMtools (87) (v1.3.1). Coverage depth across the scaffolds was calculated using the MetaBAT2 (88) (v2.12.1) jgi_summarize_bam_contig_depths script, and this information was then used by MetaBAT2 to bin the scaffolds with default parameters. The quality of the bins was assessed using the CheckM (89) (v1.0.11) lineage workflow, which included taxonomic assignment, with plots to visualize these results also produced by CheckM. To phylogenetically locate the bins, GTDBTk's classify workflow (v0.16) (https://github.com/Ecogenomics/GTDBTk) was run with default parameters. The resulting tree was then manipulated in R (v3.5.1) with the ape package (v5.1) to, for instance, find the nearest neighbors of each bin. Trees were visualized with the Interactive Tree of Life tool (v4.4.2) (90).

Genome Annotations. For *SI Appendix*, Table S6, bins of interest were annotated by RASTtk workflow (91), and subsystems were manually analyzed in the SEED viewer. For *SI Appendix*, Table S7, bins of interest were annotated with eggNOG-mapper (v1.0.3) (92), and the number of genes matching OGs of interest was counted. HMMER (v3.2.1; http://hmmer.org/) program hmmscan was used to identify and count matches for TIGRFAMs of interest using the trusted cutoff (–cut_tc) for potential hits.

Secondary Metabolite Cluster Prediction. Bioinformatic analysis of natural product genes was conducted as described previously (10). Briefly, all assembled contigs >1,500 bp were subjected to the antiSMASH standalone toolkit (v4.0.2) combined with manual Basic Local Alignment Search Tool (BLAST) analysis and conserved domain searches of uncertain regions. All manual annotation and routine bioinformatic analyses were performed using Geneious (v7.1.9) created by Biomatters (available from https://www.geneious.com/); *trans*-AT PKS gene clusters were further analyzed by TransATor (https://transator.ethz.ch/).

PCR-Based Verification of Clusters and 16S rRNA Sequences. PCRs for gap closing and gene amplification were performed with Q5 High-Fidelity DNA Polymerase (New England BioLabs). A typical PCR (25 μL) contained 1× Q5 reaction buffer, 200 μM deoxyribose nucleoside triphosphates (dNTPs), 0.5 μM of each primer (*SI Appendix*, Table S14, no. 45 to 80 for the pate-amine cluster, no. 81 to 89 for 16S rRNA genes, no. 90 to 105 for orphan cluster no. 16, no. 106 to 115 for orphan cluster no. 9, and no. 116 to 123 for orphan cluster no. 3), 20 to 50 ng template DNA, and 0.5 U Q5 High-Fidelity DNA Polymerase. The reaction was heated to 98 °C for 30 s followed by 30 cycles of 98 °C for 10 s, 62 °C for 20 s, and 72 °C for 30 s per kilobase DNA target sequence. At the end, a final incubation at 72 °C for 2 min was performed. DNA fragments were either directly sequenced or subcloned into the pCR-Blunt II-TOPO vector using the Zero Blunt TOPO PCR Cloning Kit (Invitrogen). Plasmids were transformed into *E. coli* DH5α and sequenced using the M13 forward and reverse primers.

Gene Expression and Protein Production and Purification. Expression constructs were constructed using typical PCR and restriction–endonuclease-mediated cloning techniques, fusion PCR, or Gibson assembly. A typical PCR (25 μL) contained 1× Q5 reaction buffer, 200 μM dNTPs, 0.5 μM of each primer (*SI Appendix*, Table S14, no. 1 to 30 for the gananamide cluster and no. 31 to 44 for the peloruside cluster), 20 to 50 ng template DNA, and 0.5 U Q5 High-Fidelity DNA Polymerase. The reaction was heated to 98 °C for 30 s followed by 30 cycles of 98 °C for 10 s, 62 °C for 20 s, and 72 °C for 30 s per kilobase DNA target sequence. At the end, a final incubation at 72 °C for 2 min was performed. For fusion PCR, 0.5 μL of the initial PCRs were used as a template, and only the outermost flanking primers were used. Gibson assembly was performed as per the manufacturer's instructions (Gibson Assembly Master Mix; New England BioLabs). *E. coli* BL21 (DE3) was transformed with plasmids, and expression cultures were inoculated from overnight cultures in a 1:100 (vol/vol) dilution in Terrific Broth (TB) medium. Cultures were grown (37 °C, 200 rpm) to an OD$_{600}$ of 1.2 to 1.5 and cooled on ice for 30 min. Gene expression was induced by adding 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 0.2% (wt/vol) L-arabinose, and the cultures were incubated (16 °C, 200 rpm) overnight. After harvesting the cells by centrifugation at 3,220 × *g* and 4 °C for 10 min, the pellet was resuspended in lysis buffer (50 mM NaH$_2$PO$_4$, pH 8, 300 mM NaCl, 10 mM imidazole, 10% glycerol), lysed by sonication, and centrifuged at 12,000 × *g* and 4 °C for 30 min.

Proteins were purified using the Protino Ni-NTA resin (Macherey-Nagel) according to the manufacturer's protocol.

PelA and PelB In Vitro Assays and HPLC-MS Analysis. *Nhis-pelA* in pCDFDuet-1 was expressed in *E. coli* BAP1 (93) in order to obtain activated (phosphopantetheinylated) ACPs. *Nhis-pelB* in pET28b was expressed in *E. coli* BL21 (DE3). Cultures were grown in 250-mL flasks (50 mL TB medium, 30 °C, 200 rpm) to an OD$_{600}$ of 1.2. Gene expression was induced by adding 0.5 mM IPTG. Eluted proteins were desalted using a PD MiniTrap G-25 column (GE Healthcare Biosciences) in 50 mM Tris buffer, pH 7.8, containing 100 mM NaCl, 50 mM KCl, and 5% glycerol. Incubations were performed in 50 mM Tris, pH 7.8, containing 50 mM MgCl$_2$, 1 mM carboxylic acid substrate, 1 mM adenosine triphosphate 0.5 mM coenzyme A (CoASH), and 1 μM PelA/PelB in a total reaction volume of 50 μL. After incubation at 25 °C for 3 h, 50 μL acetonitrile was added to quench the reaction. The reaction mixture was centrifuged at 20,000 × *g* for 10 min, and the supernatant was analyzed by HPLC-MS on a Phenomenex Kinetex 2.6-μm XB-C18 100-Å (150 × 2.1-mm) column for PelB assays and a Phenomenex Aeris WIDEPORE 3.6-μm C4 (50 × 2.1 mm) for assays including PelA. The column was heated to 50 °C, and the solvents used were water with 0.1% (vol/vol) formic acid (solvent A) and acetonitrile with 0.1% (vol/vol) formic acid (solvent B). For the PelB assays, a flow rate of 0.5 mL/min with solvent B at 5% from 0 to 2 min, 5 to 98% from 2 to 12 min, 98% from 12 to 15 min, 98 to 5% from 15 to 17 min, and 5% from 17 to 19 min was used. electrospray ionization mass spectrometry (ESI-MS) was performed in positive ion mode with a spray voltage of 3,500 V, a capillary temperature of 280 °C, probe heater temperature of 475 °C, and an S-Lens radio frequency (RF) level of 50. Full MS was performed at a resolution of 140,000 (automated gain control [AGC] target 1e6, maximum injection time [IT] 150 ms, range 100 to 1,000 *m/z*). For the assays including PelA, 0.1% trifluoroacetic acid was added to the solvents instead of formic acid. A flow rate of 0.2 mL/min with solvent B at 10% from 0 to 2 min, 10 to 50% from 2 to 5 min, 50 to 98% from 5 to 12 min, 98% from 12 to 18 min, 98 to 10% from 18 to 19 min, and 10% from 19 to 20 min was used. ESI-MS was performed in positive ion mode with a spray voltage of 3,500 V, a capillary temperature of 280 °C, probe heater temperature of 475 °C, and an S-Lens RF level of 100. Full MS was performed at a resolution of 140,000 (AGC target 1e6, maximum IT 150 ms, range 870 to 2,000 *m/z*).

Expression Conditions for Gananamide Proteins. *Nhis-ganA* in pCDFDuet-1 with or without *ganF* in pACYCDuet-1 or *ganD* in pBAD-Myc-HisA was expressed in *E. coli* BL21 (DE3). Cultures were grown (37 °C, 250 rpm) to an optical density at 600 nM (OD$_{600}$) of 1.6 to 2.0 and cooled on ice for 30 min. Gene expression was induced by adding 0.1 mM IPTG and 0.2% L-arabinose for epimerase induction. The cultures were then incubated at 16 °C, 200 rpm overnight. The cells were harvested by centrifugation (3,220 × *g*, 4 °C, 10 min). The cell pellets were resuspended and lysed, and the Nhis-GanA precursor was purified as described above. The elution fractions were concentrated using Amicon Ultra 0.5-mL centrifugal filters with a 3-kDa molecular weight cut-off (Merck).

ODIS Experiments. *E. coli* Tuner (DE3) was cotransformed with *Nhis-ganA* precursor in pACYCDuet-1 and *poyD* or *aerD* in pCDFBAD-Myc-HisA (66), which is derived from pBAD/Myc-His A with the native origin of replication replaced by that of pCDFDuet. Cells were plated on Luria–Bertani (LB) agar containing chloramphenicol (25 μg/mL) and ampicillin (100 μg/mL). Two separate 50-mL Falcon tubes containing TB medium (15 mL), chloramphenicol (25 μg/mL), and ampicillin (100 μg/mL) were inoculated with overnight culture in a 1:100 (vol/vol) dilution and incubated at 250 rpm and 37 °C to an OD$_{600}$ of 1.5 to 2.0. The cultures were then cooled on ice for 30 min, induced with 0.1 mM IPTG, and incubated at 250 rpm and 16 °C for 16 h. The cultures were centrifuged at 3,220 × *g* and 4 °C for 10 min, and the supernatant was removed. The cell pellets were then washed with TB medium (2 × 15 mL) to remove any residual IPTG. In the second wash, the cells were incubated at 200 rpm and 16 °C for 1 h to further metabolize intracellular IPTG. The washed cell pellet was reconstituted in TB medium in D$_2$O (15 mL) containing 0.2% (wt/vol) L-arabinose in D$_2$O and incubated at 250 rpm and 16 °C for 24 h. The cells were harvested by centrifugation at 3,220 × *g* and 4 °C for 10 min and subjected to protein purification. The elution fractions were concentrated as described above.

Proteolytic Digests and HPLC-MS/MS Analysis. Endoproteinase GluC digests (40 μL) were typically conducted using 19 μL concentrated elution fraction, 20 μL 2× GluC reaction buffer, and 1 μL endoproteinase GluC (0.2 mg/mL). The reaction mixtures were incubated at 37 °C overnight. After incubation, 40 μL acetonitrile was added to quench the reaction. The reaction mixtures were

centrifuged at 20,000 × g for 10 min, and 15 μL mixtures were analyzed by HPLC-MS on a Phenomenex Aeris WIDEPORE 3.6-μm C4 (50 × 2.1-mm) column. The column was heated to 50 °C, and the solvents used were water with 0.1% (vol/vol) formic acid (solvent A) and acetonitrile with 0.1% (vol/vol) formic acid (solvent B). A flow rate of 0.8 mL/min with solvent B at 5% from 0 to 2 min, 5 to 20% from 2 to 5 min, 20 to 65% from 5 to 15 min, 65 to 98% from 15 to 17 min, 98% from 17 to 18 min, and 98 to 5% from 18 to 20 min was used. ESI-MS was performed in positive ion mode with a spray voltage of 3,500 V, a capillary temperature of 280 °C, probe heater temperature of 475 °C, and an S-Lens RF level of 100. Full MS was performed at a resolution of 35,000 (AGC target 1e6, maximum IT 250 ms, range 300 to 2,000 m/z). Parallel reaction monitoring (PRM) or data-dependent tandem mass spectrometry (MSMS) was performed at a resolution of 17,500 (AGC target between 2e5 and 1e6, maximum IT between 100 and 500 ms, isolation windows in the range of 1.1 to 2.2 m/z) using a stepped normalized collision energy (NCE) of 18, 20, and 22. Scan ranges, inclusion lists, charge exclusions, and dynamic exclusions were adjusted as needed.

For proteinase K digests (40 μL), 16 μL concentrated elution fraction was added to 20 μL 2× proteinase K buffer (50 mM Tris, pH 8.0, 2 mM CaCl₂, final concentration) and 4 μL proteinase K (0.2 mg/mL). The reaction mixtures were incubated at 50 °C for 12 h and quenched by addition of 40 μL acetonitrile, and 15 μL of the mixture was analyzed by HPLC-MS on a Phenomenex Kinetex 2.6-μm XB-C18 100-Å (150 × 2.1-mm) column. The column was heated to 50 °C, and the solvents used were water with 0.1% (vol/vol) formic acid (solvent A) and acetonitrile with 0.1% (vol/vol) formic acid (solvent B). A flow rate of 0.5 mL/min with solvent B at 5% from 0 to 2 min, 5 to 50% from 2 to 5 min, 50 to 65% from 5 to 15 min, 65 to 98% from 15 to 17 min, 98% from 17 to 18 min, and 98 to 5% from 18 to 20 min was used. ESI-MS was performed in positive ion mode with a spray voltage of 3,500 V, a capillary temperature of 280 °C, probe heater temperature of 475 °C, and an S-Lens RF level of 100. Full MS was performed at a resolution of 35,000 (AGC target 1e6, maximum IT 250 ms, range 300 to 2,000 m/z). PRM or data-dependent MSMS was performed at a resolution of 17,500 (AGC target between 2e5 and 1e6, maximum IT between 100 and 500 ms, isolation windows in the range of 1.1 to 2.2 m/z) using a stepped NCE of 18, 20, and 22. Scan ranges, inclusion lists, charge exclusions, and dynamic exclusions were adjusted as needed.

MS/MS Analysis–MaxQuant. For MS-based identification of proteinase K-treated peptide fragments, the program MaxQuant (94) (v1.5.2.8) was used. The following parameters were changed from default settings in a typical run: variable modifications, as needed with customized masses made in the Andromeda configuration tab; digestion mode, unspecific; maximum peptide mass (daltons), 10,000; minimum peptide length for unspecific search, 4; and maximum peptide length for unspecific search, 50. On completion of the run, the evidence text file was used to map candidate peptides with MSMS scan numbers. The fragments were manually verified and annotated using the Xcalibur Qual Browser software (Thermo Fisher Scientific).

1. J. L. Cleary, A. R. Condren, K. E. Zink, L. M. Sanchez, Calling all hosts: Bacterial communication in situ. Chem 2, 334–358 (2017).
2. M. R. Wilson, L. Zha, E. P. Balskus, Natural product discovery from the human microbiome. J. Biol. Chem. 292, 8546–8552 (2017).
3. L. V. Flórez, P. H. W. Biedermann, T. Engl, M. Kaltenpoth, Defensive symbioses of animals with prokaryotic and eukaryotic microorganisms. Nat. Prod. Rep. 32, 904–936 (2015).
4. J. A. Zumberge et al., Demosponge steroid biomarker 26-methylstigmastane provides evidence for Neoproterozoic animals. Nat. Ecol. Evol. 2, 1709–1714 (2018).
5. M. F. Mehbub, J. Lei, C. Franco, W. Zhang, Marine sponge derived natural products between 2001 and 2010: Trends and opportunities for discovery of bioactives. Mar. Drugs 12, 4539–4577 (2014).
6. U. Hentschel, J. Piel, S. M. Degnan, M. W. Taylor, Genomic insights into the marine sponge microbiome. Nat. Rev. Microbiol. 10, 641–654 (2012).
7. J. Tout et al., Redefining the sponge-symbiont acquisition paradigm: Sponge microbes exhibit chemotaxis towards host-derived compounds. Environ. Microbiol. Rep. 9, 750–755 (2017).
8. M. W. Taylor et al., 'Sponge-specific' bacteria are widespread (but rare) in diverse marine environments. ISME J. 7, 438–443 (2013).
9. N. S. Webster, T. Thomas, The sponge hologenome. MBio 7, e00135-16 (2016).
10. M. C. Wilson et al., An environmental bacterial taxon with a large and distinct metabolic repertoire. Nature 506, 58–62 (2014).
11. C. J. Wegerski, J. Hammond, K. Tenney, T. Matainaho, P. Crews, A serendipitous discovery of isomotuporin-containing sponge populations of Theonella swinhoei. J. Nat. Prod. 70, 89–94 (2007).
12. G. Lackner, E. E. Peters, E. J. Helfrich, J. Piel, Insights into the lifestyle of uncultured bacterial natural product factories associated with marine sponges. Proc. Natl. Acad. Sci. U.S.A. 114, E347–E356 (2017).
13. M. J. Helf, A. Jud, J. Piel, Enzyme from an uncultivated sponge bacterium catalyzes S-methylation in a ribosomal peptide. ChemBioChem 18, 444–450 (2017).
14. T. Mori et al., Single-bacterial genomics validates rich and varied specialized metabolism of uncultivated Entotheonella sponge symbionts. Proc. Natl. Acad. Sci. U.S.A. 115, 1718–1723 (2018).
15. T. Wakimoto et al., Calyculin biogenesis from a pyrophosphate protoxin produced by a sponge symbiont. Nat. Chem. Biol. 10, 648–655 (2014).
16. Y. Nakashima, Y. Egami, M. Kimura, T. Wakimoto, I. Abe, Metagenomic analysis of the sponge Discodermia reveals the production of the yyanobacterial natural product kasumigamide by 'Entotheonella.' PLoS One 11, e0164468 (2016).
17. E. W. Schmidt, A. Y. Obraztsova, S. K. Davidson, D. J. Faulkner, M. G. Haygood, Identification of the antifungal peptide-containing symbiont of the marine sponge Theonella swinhoei as a novel δ-proteobacterium, "Candidatus Entotheonella palauensis. Mar. Biol. 136, 969–977 (2000).
18. M. D. Unson, N. D. Holland, D. J. Faulkner, 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, 1–11 (1994).
19. P. M. Flatt et al., 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 (2005).
20. V. Agarwal et al., Metagenomic discovery of polybrominated diphenyl ether biosynthesis by marine sponges. Nat. Chem. Biol. 13, 537–543 (2017).
21. W. H. Gerwick, B. S. Moore, Lessons from the past and charting the future of marine natural products drug discovery and chemical biology. Chem. Biol. 19, 85–98 (2012).
22. L. J. Habener, J. N. A. Hooper, A. R. Carroll, Chemical and biological aspects of marine sponges from the family Mycalidae. Planta Med. 82, 816–831 (2016).
23. P. R. Bergquist, P. J. Fromont, "The marine fauna of New Zealand: Porifera, demospongiae, part 4 (Poecilosclerida)" in New Zealand Oceanographic Institute Memoir 96 (New Zealand Oceanographic Institute, Wellington, New Zealand, 1988), pp. 1–197.
24. S. A. Dyshlovoy et al., Mycalamide A shows cytotoxic properties and prevents EGF-induced neoplastic transformation through inhibition of nuclear factors. Mar. Drugs 10, 1212–1224 (2012).
25. N. B. Perry, J. W. Blunt, M. H. G. Munro, L. K. Pannell, A. Mycalamide, An antiviral compound from a New Zealand sponge of the genus Mycale. J. Am. Chem. Soc. 110, 4850–4851 (1988).
26. W.-K. Low et al., Inhibition of eukaryotic translation initiation by the marine natural product pateamine A. Mol. Cell 20, 709–722 (2005).
27. P. T. Northcote, J. W. Blunt, M. H. G. Munro, Pateamine: A potent cytotoxin from the New Zealand marine sponge, Mycale sp. Tetrahedron Lett. 32, 6411–6414 (1991).
28. K. A. Hood et al., Peloruside A, a novel antimitotic agent with paclitaxel-like microtubule- stabilizing activity. Cancer Res. 62, 3356–3360 (2002).
29. L. M. West, P. T. Northcote, C. N. Battershill, Peloruside A: A potent cytotoxic macrolide isolated from the New Zealand marine sponge Mycale sp. J. Org. Chem. 65, 445–449 (2000).
30. T. N. Gaitanos et al., Peloruside A does not bind to the taxoid site on beta-tubulin and retains its activity in multidrug-resistant cell lines. Cancer Res. 64, 5063–5067 (2004).
31. M. J. Page, S. J. Handley, P. T. Northcote, D. Cairney, R. C. Willan, Successes and pitfalls of the aquaculture of the sponge Mycale hentscheli. Aquaculture 312, 52–61 (2011).
32. M. Page, L. West, P. Northcote, C. Battershill, M. Kelly, Spatial and temporal variability of cytotoxic metabolites in populations of the New Zealand sponge Mycale hentscheli. J. Chem. Ecol. 31, 1161–1174 (2005).
33. J. Piel, A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles. Proc. Natl. Acad. Sci. U.S.A. 99, 14002–14007 (2002).
34. A. Nakabachi et al., Defensive bacteriome symbiont with a drastically reduced genome. Curr. Biol. 23, 1478–1484 (2013).

35. A. Kampa et al., Metagenomic natural product discovery in lichen provides evidence for a family of biosynthetic pathways in diverse symbioses. Proc. Natl. Acad. Sci. U.S.A. 110, E3129–E3137 (2013).

36. J. Piel et al., Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge Theonella swinhoei. Proc. Natl. Acad. Sci. U.S.A. 101, 16222–16227 (2004).

37. K. M. Fisch et al., Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. Nat. Chem. Biol. 5, 494–501 (2009).

38. C. Schleissner et al., Bacterial production of a pederin analogue by a free-living marine Alphaproteobacterium. J. Nat. Prod. 80, 2170–2173 (2017).

39. A. Kust et al., Discovery of a pederin family compound in a nonsymbiotic bloomforming Cyanobacterium. ACS Chem. Biol. 13, 1123–1129 (2018).

40. J. Piel, G. Wen, M. Platzer, D. Hui, Unprecedented diversity of catalytic domains in the first four modules of the putative pederin polyketide synthase. ChemBioChem 5, 93–98 (2004).

41. E. J. Helfrich, J. Piel, Biosynthesis of polyketides by trans-AT polyketide synthases. Nat. Prod. Rep. 33, 231–316 (2016).

42. C. T. Calderone, Isoprenoid-like alkylations in polyketide biosynthesis. Nat. Prod. Rep. 25, 845–853 (2008).

43. S. Nurk, D. Meleshko, A. Korobeynikov, P. A. Pevzner, metaSPAdes: A new versatile metagenomic assembler. Genome Res. 27, 824–834 (2017).

44. K. Blin et al., antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. Nucleic Acids Res. 45, W36–W41 (2017).

45. T. Nguyen et al., Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. Nat. Biotechnol. 26, 225–233 (2008).

46. T. Thomas et al., Diversity, structure and convergent evolution of the global sponge microbiome. Nat. Commun. 7, 11870 (2016).

47. U. Hentschel et al., Molecular evidence for a uniform microbial community in sponges from different oceans. Appl. Environ. Microbiol. 68, 4431–4440 (2002).

48. M. D. Tianero, J. N. Balaich, M. S. Donia, Localized production of defence chemicals by intracellular symbionts of Haliclona sponges. Nat. Microbiol. 4, 1149–1159 (2019).

49. J. C. Kwan et al., Genome streamlining and chemical defense in a coral reef symbiosis. Proc. Natl. Acad. Sci. U.S.A. 109, 20655–20660 (2012).

50. R. A. Meoded et al., A polyketide synthase component for oxygen insertion into polyketide backbones. Angew. Chem. Int. Ed. Engl. 57, 11644–11648 (2018).

51. C. Quast et al., The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D596 (2013).

52. R. M. Tian et al., The deep-sea glass sponge Lophophysema eversa harbours potential symbionts responsible for the nutrient conversions of carbon, nitrogen and sulfur. Environ. Microbiol. 18, 2481–2494 (2016).

53. E. J. N. Helfrich et al., Automated structure prediction of trans-acyltransferase polyketide synthase products. Nat. Chem. Biol. 15, 813–821 (2019).

54. K. Bloudoff, C. D. Fage, M. A. Marahiel, T. M. Schmeing, Structural and mutational analysis of the nonribosomal peptide synthetase heterocyclization domain provides insight into catalysis. Proc. Natl. Acad. Sci. U.S.A. 114, 95–100 (2017).

55. J. B. Biggins, M. A. Ternei, S. F. Brady, Malleilactone, a polyketide synthase-derived virulence factor encoded by the cryptic secondary metabolome of Burkholderia pseudomallei group pathogens. J. Am. Chem. Soc. 134, 13192–13195 (2012).

56. J. Franke, K. Ishida, C. Hertweck, Genomics-driven discovery of burkholderic acid, a noncanonical, cryptic polyketide from human pathogenic Burkholderia species. Angew. Chem. Int. Ed. Engl. 51, 11611–11615 (2012).

57. S. Spring et al., Characterization of the first cultured representative of Verrucomicrobia subdivision 5 indicates the proposal of a novel phylum. ISME J. 10, 2801–2816 (2016).

58. D. H. Parks et al., A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. 36, 996–1004 (2018).

59. A. S. Eustáquio, J. E. Janso, A. S. Ratnayake, C. J. O'Donnell, F. E. Koehn, Spliceostatin hemiketal biosynthesis in Burkholderia spp. is catalyzed by an iron/α-ketoglutarate-dependent dioxygenase. Proc. Natl. Acad. Sci. U.S.A. 111, E3376–E3385 (2014).

60. X. Liu et al., Genomics-guided discovery of thailanstatins A, B, and C As pre-mRNA splicing inhibitors and antiproliferative agents from Burkholderia thailandensis MSMB43. J. Nat. Prod. 76, 685–693 (2013).

61. G. Pan et al., Discovery of the leinamycin family of natural products by mining actinobacterial genomes. Proc. Natl. Acad. Sci. U.S.A. 114, E11131–E11140 (2017).

62. M. F. Freeman et al., Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides. Science 338, 387–390 (2012).

63. T. Hamada et al., Solution structure of polytheonamide B, a highly cytotoxic nonribosomal polypeptide from marine sponge. J. Am. Chem. Soc. 132, 12941–12945 (2010).

64. M. F. Freeman, M. J. Helf, A. Bhushan, B. I. Morinaka, J. Piel, Seven enzymes create extraordinary molecular complexity in an uncultivated bacterium. Nat. Chem. 9, 387–395 (2017).

65. B. I. Morinaka et al., Radical S-adenosyl methionine epimerases: Regioselective introduction of diverse D-amino acid patterns into peptide natural products. Angew. Chem. Int. Ed. Engl. 53, 8503–8507 (2014).

66. A. Bhushan, P. J. Egli, E. E. Peters, M. F. Freeman, J. Piel, Genome mining- and synthetic biology-enabled production of hypermodified peptides. Nat. Chem. 11, 931–939 (2019).

67. B. I. Morinaka, M. Verest, M. F. Freeman, M. Gugger, J. Piel, An orthogonal D$_2$O-based induction system that provides insights into D-amino acid pattern formation by radical S-adenosylmethionine peptide epimerases. Angew. Chem. Int. Ed. Engl. 56, 762–766 (2017).

68. A. J. Pinto et al., Metagenomic evidence for the presence of comammox Nitrospira-like bacteria in a drinking water system. MSphere 1, e00054-15 (2015).

69. R. W. M. Van Soest et al., Global diversity of sponges (Porifera). PLoS One 7, e35105 (2012).

70. L. Pita, L. Rix, B. M. Slaby, A. Franke, U. Hentschel, The sponge holobiont in a changing ocean: From microbes to ecosystems. Microbiome 6, 46 (2018).

71. A. Kanakkanthara, P. T. Northcote, J. H. Miller, Peloruside A: A lead non-taxoid-site microtubule-stabilizing agent with potential activity against cancer, neurodegeneration, and autoimmune disease. Nat. Prod. Rep. 33, 549–561 (2016).

72. M. S. Donia, W. F. Fricke, J. Ravel, E. W. Schmidt, Variation in tropical reef symbiont metagenomes defined by secondary metabolism. PLoS One 6, e17897 (2011).

73. T. E. Smith et al., Accessing chemical diversity from the uncultivated symbionts of small marine animals. Nat. Chem. Biol. 14, 179–185 (2018).

74. R. Ueoka et al., Metabolic and evolutionary origin of actin-binding polyketides from diverse organisms. Nat. Chem. Biol. 11, 705–712 (2015).

75. Z. Lin, J. P. Torres, M. D. Tianero, J. C. Kwan, E. W. Schmidt, Origin of chemical diversity in Prochloron-tunicate symbiosis. Appl. Environ. Microbiol. 82, 3450–3460 (2016).

76. R. L. L. Kellner, K. Dettner, Differential efficacy of toxic pederin in deterring potential arthropod predators of Paederus (Coleoptera: Staphylinidae) offspring. Oecologia 107, 293–300 (1996).

77. T. Yamada, M. Hamada, P. Floreancig, A. Nakabachi, Diaphorin, a polyketide synthesized by an intracellular symbiont of the Asian citrus psyllid, is potentially harmful for biological control agents. PLoS One 14, e0216319 (2019).

78. M. C. Wilson, J. Piel, Metagenomic approaches for exploiting uncultured bacteria as a resource for novel biosynthetic enzymology. Chem. Biol. 20, 636–647 (2013).

79. A. Crits-Christoph, S. Diamond, C. N. Butterfield, B. C. Thomas, J. F. Banfield, Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. Nature 558, 440–444 (2018).

80. G. E. Lim, M. G. Haygood, "Candidatus Endobugula glebosa," a specific bacterial symbiont of the marine bryozoan Bugula simplex. Appl. Environ. Microbiol. 70, 4921–4929 (2004).

81. J. Lopera, I. J. Miller, K. L. McPhail, J. C. Kwan, Increased biosynthetic gene dosage in a genome-reduced defensive bacterial symbiont. mSystems 2, e00096-17 (2017).

82. J. Zan et al., A microbial factory for defensive kahalalides in a tripartite marine symbiosis. Science 364, eaaw6732 (2019).

83. S. A. Anderson, P. T. Northcote, M. J. Page, Spatial and temporal variability of the bacterial community in different chemotypes of the New Zealand marine sponge Mycale hentschell. FEMS Microbiol. Ecol. 72, 328–342 (2010).

84. C. Gurgui, J. Piel, Metagenomic approaches to identify and isolate bioactive natural products from microbiota of marine sponges. Methods Mol. Biol. 668, 247–264 (2010).

85. D. Li, C. M. Liu, R. Luo, K. Sadakane, T. W. Lam, MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31, 1674–1676 (2015).

86. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009).

87. H. Li et al.; 1000 Genome Project Data Processing Subgroup, The sequence alignment/ map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).

88. D. D. Kang, J. Froula, R. Egan, Z. Wang, MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3, e1165 (2015).

89. D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz, G. W. Tyson, CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–1055 (2015).

90. I. Letunic, P. Bork, Interactive tree of life (iTOL) v4: Recent updates and new developments. Nucleic Acids Res. 47, W256–W259 (2019).

91. T. Brettin et al., RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci. Rep. 5, 8365 (2015).

92. J. Huerta-Cepas et al., Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol. Biol. Evol. 34, 2115–2122 (2017).

93. B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane, C. Khosla, Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. Science 291, 1790–1792 (2001).

94. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. 26, 1367–1372 (2008).