

Chemical richness and diversity of uncultivated ‘*Entotheonella*’ symbionts in marine sponges

Received: 10 September 2024

Accepted: 6 October 2025

Published online: 13 November 2025



Maria Dell ^{1,17}, Masato Kogawa^{2,3,4,17}, Alena B. Streiff ^{1,17}, Taro Shiraishi ^{1,5,6}, Alessandro Lotti¹, Christoph M. Meier ¹, Michelle A. Schorn ⁷, Christopher Field¹, Jackson K. B. Cahn¹, Hiromi Yokoyama ⁸, Yuito Yamada⁸, Eike Peters ¹, Yoko Egami⁸, Yu Nakashima ⁹, Karen Co Tan ⁹, Christian Rückert ¹⁰, Mohammad Alanjary ¹¹, Jörn Kalinowski¹⁰, Tomohisa Kuzuyama ^{5,6}, Paco Cardenas ^{12,13}, Shirley Pomponi^{14,15}, Detmer Sipkema⁷, Amy Wright ¹⁵, Kentaro Takada¹⁶, Ikuro Abe ⁹, Toshiyuki Wakimoto ^{8,18} , Haruko Takeyama ^{2,3,4,18}  & Jörn Piel ^{1,18} 

Marine sponges are the source of numerous bioactive natural products that serve as chemical defenses and provide pharmaceutical leads for drug development. For some of the compounds, symbiotic bacteria have been established as the actual producers. Among the known sponge symbionts, ‘*Candidatus Entotheonella*’ members stand out because of their abundant and variable biosynthetic gene clusters (BGCs). Here, to obtain broader insights into this producer taxon, we conduct a comparative analysis on eight sponges through metagenomic and single-bacterial sequencing and biochemical studies. The data suggest sets of biosynthetic genes that are largely unique in 14 ‘*Entotheonella*’ candidate species and a member of a sister lineage named ‘*Candidatus Proxinella*’. Four biosynthetic loci were linked in silico or experimentally to cytotoxins, antibiotics and the terpene cembrene A from corals. The results support widespread and diverse bacterial roles in the chemistry of sponges and aid the development of sustainable production methods for sponge-derived therapeutics.

Sponges (Porifera) are ancient metazoans with unusually diverse bioactive natural products (NPs)^{1,2} with suspected or demonstrated roles as chemical defenses against grazers or epibionts^{3–8}. In addition, sponge NPs are a rich resource for drug development, with spongouridine and halichondrins as examples of important leads in antiviral and cancer therapy^{9–11}. While some sponge NPs are host synthesized^{12,13}, evidence increases that many others are products of the sponge microbiome^{14–17}. Studying these often diverse bacterial communities remains challenging, as most members resist cultivation, making functional characterization difficult¹⁸. Bacterial origins have been established for several sponge NPs, including the anticancer drug candidates psymberin¹⁹,

peloruside A^{20,21} and renieramycin²². The most prolific known sponge symbionts belong to the candidate genus ‘*Entotheonella*’ (quote format refers to uncultivated status), first reported by Bewley, Schmidt, Haygood, Faulkner and coworkers from a Palauan *Theonella swinhoei* sponge^{23–26}. *T. swinhoei* displays remarkable chemical diversity across distinct chemotypes^{24,27,28}. In the Palauan variant, chemical analysis localized theopalauamide (**1**) (Fig. 1) to a cell fraction enriched in filamentous bacteria²⁶ named ‘*Candidatus Entotheonella palauensis*’, suggesting it as the producer²³. Further genomic and biosynthetic work identified ‘*Entotheonella*’ producers in other *T. swinhoei* chemotypes, assigning them to four candidate species^{29–31}.

A full list of affiliations appears at the end of the paper. ✉ e-mail: wakimoto@pharm.hokudai.ac.jp; haruko-takeyama@waseda.jp; jpiel@ethz.ch

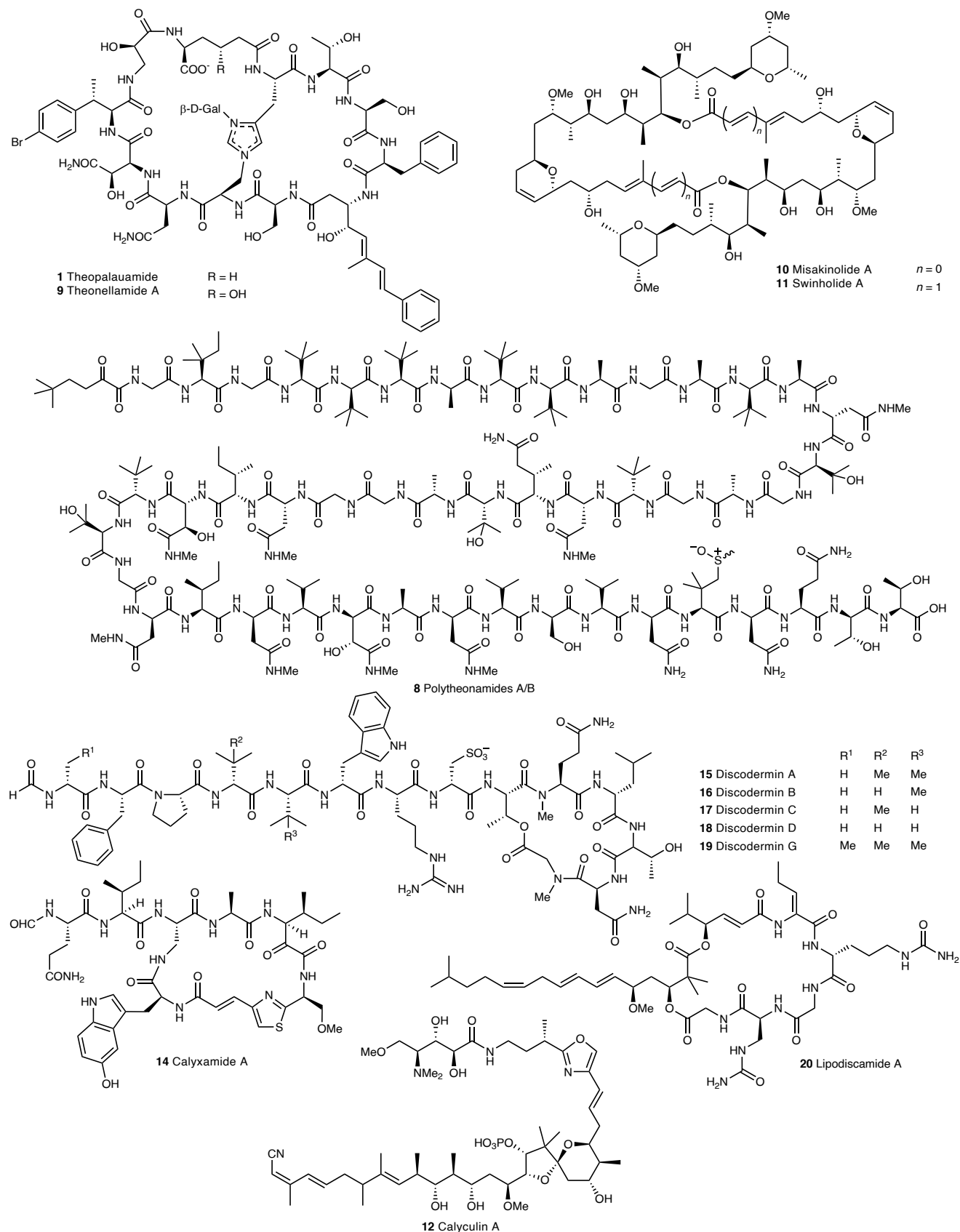


Fig. 1 | Sponge-derived NPs relevant to this study (selection). A full overview of numbered compounds is provided in Supplementary Fig. 1.

Table 1 | Genome-sequenced ‘Entotheonella’ phylotypes and their host sponges analyzed in previous work or the current study (specified in column 1). Further details can be found in Supplementary Table 1

Sponge (suborder)	Known compounds (numbers in parentheses refer to representative structures in Fig. 1 and Supplementary Fig. 1)	Symbiont variants (E. refers to ‘Entotheonella’, P. to ‘Proxinnella’)
<i>T. swinhoei</i> Y (Astrophorina); previous studies	Konbamide (2), keramamides (3), nazumamide A (4), pseudotheonamides (5), cyclotheonamide (6), onnamides and theopederins (7), polytheonamides (8), aurantosides and sterols	‘Ca.E.factor’ TSY1 ‘Ca.E.gemina’ TSY2 ‘Ca.E.arcus’ YB-1
<i>T. swinhoei</i> YB (Astrophorina); this study	See above	‘Ca.E.factor’ TSYB1 ‘Ca.E.gemina’ TSYB2 ‘Ca.E.mitsugo’ TSYB3
<i>T. swinhoei</i> WA (Astrophorina); previous study	Theonellamides (9) and misakinolides (10)	‘Ca.E.serta’ TSWA1
<i>T. swinhoei</i> WB (Astrophorina); this study	Theonellamides (9) and swinholides (11)	‘Ca.E.serta’ TSWB1 ‘Ca.E.consors’ TSWB2
<i>Theonella</i> sp. 1 BA (Astrophorina); this study	Unknown	‘Ca.E.melakyensis’ TCBA1 ‘Ca.E.serta’ TCBA2
<i>Theonella</i> sp. 2 BT (Astrophorina); this study	Unknown	‘Ca.E.symbiotica’ BT01 ‘Ca.E.inquilina’ BT02 ‘Ca.E.melakyensis’ BT03 ‘Ca.E.catenata’ BT04
<i>D. calyx</i> (Astrophorina); this study	Calyculins (12), kasumigamide (13), calyxamides (14), indole derivatives and cyclodipeptides	‘Ca.E.armillaria’ DC1
<i>D. kiiensis</i> (Astrophorina); this study	Discodermisins (15–19), lipodiscamides (20)	‘Ca.E.monilis’ DK1
<i>D. dissoluta</i> (Astrophorina); this study	Discodermolides	‘Ca.E.tacita’ DD1 ‘Ca.E.baccata’ DD2 ‘Ca.E.tertia’ DD3
<i>A. cribrophora</i> (Spirophorina); this study	Unknown	‘Ca.P.opulenta’ AC1

These four symbionts, members of the candidate phylum ‘Tectomicrobia’ (‘Entotheonellaeota’), feature unusually large ~10-Mb genomes with diverse biosynthetic gene clusters (BGCs) for known sponge compounds and predicted unknown NPs. In the chemically rich yellow *T. swinhoei* chemotype (Y) from Japan, ‘*Candidatus* Entotheonella factor’ produces most known polyketides and peptides, including **2–8** (Table 1, Fig. 1 and Supplementary Fig. 1)^{29,32,33}, while the coinhabiting ‘*Candidatus* Entotheonella gemina’ contains only orphan BGCs²⁹. Recently, ‘*Candidatus* Entotheonella arcus’ was found colonizing some yellow *T. swinhoei* specimens³¹. In contrast, white *T. swinhoei* chemotypes (W) from Japan and Israel contain ‘*Candidatus* Entotheonella sarta’, producing compounds such as **9–11**, in addition to containing many orphan BGCs (Table 1 and Fig. 1)^{30,34,35}.

Research on other sponge microbiomes revealed BGCs assigned to ‘Entotheonella’ by in situ hybridization. These encode the biosynthesis of calyculins (**12**)³⁶ and kasumigamides (**13**)³⁷ from *Discodermia calyx* and psymberin from a *Psammocinia* sp. sponge³⁸. However, their metabolic diversity and phylogenetic affiliation remained unknown without further genomic data. A 16S ribosomal RNA (rRNA) gene-based study suggested ‘Entotheonella’ as a diverse lineage with numerous members primarily in sponges but also detected in sediments and soil³⁹. These data and the high BGC diversity in few genome-sequenced representatives suggest a major untapped NP resource.

Here, we perform metagenomic, single-bacterial and functional studies to investigate these uncultivated organisms more broadly, particularly evaluating whether chemical richness is a general feature of this taxon. We interrogated how this feature distributes among taxon members and whether additional bioactive compounds are produced by these symbionts. Our data cover 15 candidate species across eight sponge chemotypes, assigned to 14 ‘Entotheonella’ phylotypes and an unexpected BGC-rich sister candidate genus. Results indicate high BGC diversity among ‘Entotheonella’ phylotypes with biochemically supported roles in producing both known and orphan sponge metabolites. This widespread chemical richness provides a foundation for targeted NP discovery from microbial dark matter.

Results

Selection of sponges for sequencing

We initiated our study by selecting a taxonomically and geographically diverse set of ‘Entotheonella’-containing sponges (Table 1). These comprised Japanese *Discodermia kiiensis*^{29,39}, previously identified as a source for discodermisins³⁶ and lipodiscamide cytotoxins^{40,41}, *D. calyx*³⁶, harboring the cytotoxic calyculins (for example, **12**)⁴² and calyxamides (for example, **14**), and *Discodermia dissoluta*⁴³ from the Bahamas, containing the anticancer discodermolides^{44,45}. All sponges were known to contain ‘Entotheonella’^{29,36,39,43} but lacked genome sequences. In addition, microscopy revealed symbionts with an ‘Entotheonella’-like morphology in two unidentified *Theonella* specimens with a new, blue phenotype from Japan and the Mozambique Channel. Furthermore, ‘Entotheonella’-like DNA contigs were detected in a sequenced *Aciculites cribrophora* metagenome. These three sponges with uncharacterized chemistry were also included in our study. Lastly, we reassessed the previously analyzed³⁹ chemically complex Japanese *T. swinhoei* Y with new assembly methods to generate improved ‘Entotheonella’ genomes. For this purpose, a further specimen of this chemotype, named *T. swinhoei* Y2, was collected. In total, the analyzed sponges encompassed eight specimens collected at seven locations and belonging to two suborders, at least six species and eight chemotypes (Fig. 2a).

Identification of 14 ‘Entotheonella’ phylotypes

We previously observed that some ‘Entotheonella’ variants resisted metagenomic sequencing but were amenable to single-bacterial sequencing³⁰. We, therefore, used either metagenomic or single-filament sequencing or both (Supplementary Tables 2 and 3), depending on method success. Metagenomic sequencing was used for *D. dissoluta* and *Theonella* sp. 1 BA after mechanical enrichment of filamentous bacteria. *A. cribrophora* underwent full metagenome sequencing with subsequent binning. For remaining sponges, single-bacterial sequencing was performed in addition to or instead of metagenomics using cell separation, microdroplet encapsulation, microscopy-aided

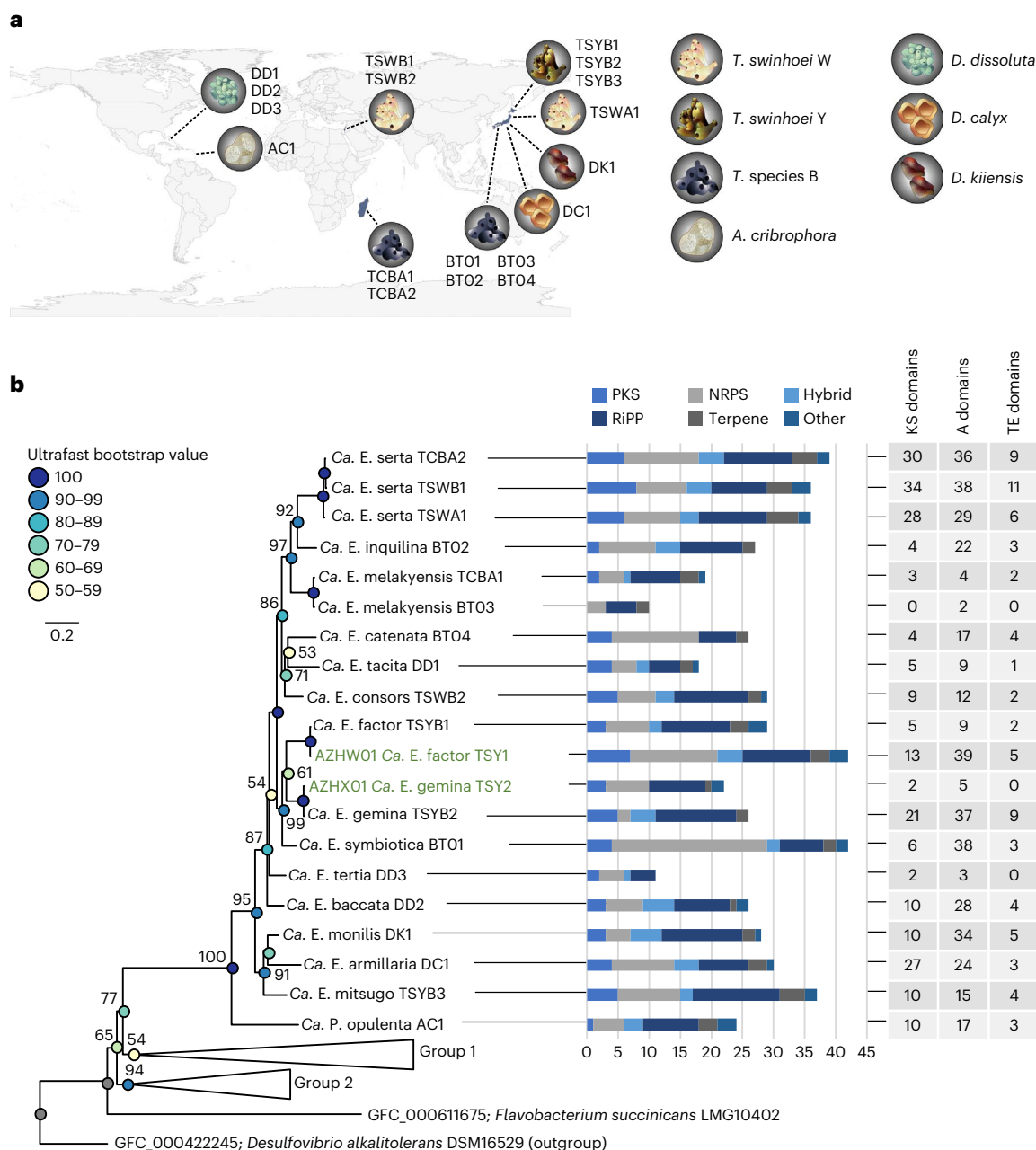


Fig. 2 | Distribution, phylogenetic relationships and metabolic potential of the analyzed sponge symbionts. a, Geographic origin of the analyzed sponges. Sponges are shown as idealized icons explained on the right. On the map, the labels besides the icons refer to the identifiers of symbionts identified in the sponges. **b**, Whole-genome phylogram generated with autoMLST of analyzed symbiont genomes and the two already published symbionts ‘*Ca. E. factor*’ TSY1 (AZHW01) and ‘*Ca. E. gemina*’ TSY2 (AZHX01), with the two latter labeled in green. Colored circles on nodes represent Ultrafast Bootstrap values. The composition of collapsed clades group 1 and group 2 is provided in the Supplementary Information. For each branch, the bar chart provides numbers

of BGCs or BGC fragments identified by antiSMASH. The BGC classes used by antiSMASH were implemented here, with minor classes being combined as ‘other’. Hybrid refers to BGC loci combining features of several classes (PKS–NRPS, NRPS–RiPPs, etc.). The rightmost table shows total counts of core domains in PKS (ketosynthase) domains and NRPS (A domains) systems, as well as termination TE domains for both systems, per genome. Domain numbers might be underrepresented for genomes with lower coverage (for example, BT02, BT04 and DD3; <70%) or overrepresented for genomes with higher contamination (for example, TSYB1, TSWA1, TCBA1 and DD1; >10%). More detailed genome statistics are provided in Supplementary Table 3.

sorting and genome amplification⁴⁶. This procedure proved valuable for *D. calyx*, where multiple metagenomic attempts failed or yielded poor genome coverage⁴⁶. Single-bacterial sequencing was also applied when plasmids or multiple ‘*Entotheonella*’ phylotypes per sponge were detected, as in *T. swinhoei* Y containing two ‘*Entotheonella*’ symbionts and one or more plasmids²⁹.

Assembly quality assessed using CheckM⁴⁷ indicated ~13% to >90% genome completeness (Supplementary Table 3). The most complete

genome was obtained for ‘*E. sarta*’ from *T. swinhoei* WB (95.7% completeness, 7.86% contamination), while the lowest-quality dataset was a metagenome-assembled genome (MAG) of ‘*Entotheonella tertia*’ from *D. dissoluta* (12.9% completeness, 0.0% contamination, 2.13-Mbp assembly size). Estimated genome sizes ranged from 5.36 to 16.54 Mbp (Supplementary Table 3), with high-quality values around 9 Mbp. Except for ‘*Poriflexus aureus*’ (~14 Mbp), previously identified in two *Theonella* sponges⁴⁶, ‘*Entotheonella*’ members feature, on the basis of

a previous large-scale study⁴⁸, some of the largest genomes identified among sponge symbionts. Phylogenomic relationships were studied using FastANI⁴⁹ and autoMLST⁵⁰. According to binning, single-bacterial sequencing and phylogenomic data (Fig. 2b), sponges contained one to four ‘Entotheonella’ phylotypes representing different candidate species. Additionally, we identified an *A. cribrophora* symbiont initially suggested by GTDB-Tk⁵¹ to belong to ‘Entotheonellaceae’ (Supplementary Information). Deeper analysis using average nucleotide identity (ANI), multilocus sequence typing (MLST) and 16S rRNA gene sequences supported its affiliation with a distinct tectomicrobial candidate genus (Supplementary Fig. 2). We included this organism, named ‘Proxinella opulenta’ AC1, in the current study because of its ‘Entotheonella’-like BGC richness, as discussed below. This contrasts with the reported genomes from the tectomicrobial genus ‘Bathynella’ with low BGC numbers³⁹.

Among all analyzed sponges, we identified 14 distinct ‘Entotheonella’ variants outside of ‘*P. opulenta*’ AC1, with proposed names in Table 1. Different sponges mostly harbored distinct ‘Entotheonella’ phylotypes, consistent with previous 16S rRNA gene-based observations³⁹, but some closely related symbionts appeared in different sponge species, suggesting horizontal transfer or inheritance from a common ancestor. For example, ‘*E. sarta*’ was identified in *T. swinhoei* WA and WB and in *Theonella* sp. 1 BA from three different locations. ‘*E. melakyensis*’ was found in the blue *Theonella* sp. sponges from Japan and the Mozambique Channel. Despite ANI values slightly below the species cutoff (93.63%; Supplementary Fig. 3), multilocus phylogeny suggests the same candidate species (Fig. 2b). Symbiont variability also existed among specimens of the same host type; *T. swinhoei* Y1 and Y2 both contain ‘*E. factor*’ and ‘*E. gemina*’, while only Y2 additionally harbors a third ‘Entotheonella’ phylotype, ‘*E. mitsugo*’. Similarly, ‘*E. sarta*’ was the sole variant in *T. swinhoei* WA but accompanied by ‘*E. consors*’ in *T. swinhoei* WB. These results reveal complex symbiont coevolution and horizontal acquisition patterns with likely consequences for sponge chemistry. Comparing our 14 ‘Entotheonella’ variants to ‘*E. arcus*’³¹ and ‘*E. halido*’⁵² reported during the completion of our study using FastANI (Supplementary Fig. 4) confirmed them as distinct candidate species. By uncovering 11 additional ‘Entotheonella’ and one ‘Proxinella’ phylotypes, we expanded knowledge beyond *T. swinhoei* sponges and identified ‘*E. mitsugo*’ as yet another phylotype in addition to ‘*E. factor*’, ‘*E. gemina*’ and ‘*E. arcus*’ in some yellow specimens of *T. swinhoei*.

We analyzed 16S rRNA gene sequences for comparison to the whole-genome tree (Fig. 2b). Using ssu-finder in CheckM⁴⁷, we identified 27 16S rRNA genes or fragments in all genomes, except for ‘*E. melakyensis*’ TCBA1, ‘*E. sarta*’ TCBA2 and ‘*E. tertia*’ DD3. Four sequences were excluded as chimeric or duplicated. Of the remaining 23, eight were of sufficient length for phylogenetic analysis. The 16S rRNA phylogram largely mirrored the whole-genome phylogeny (Supplementary Fig. 5a). To relate our variants to the theopalauamide-containing, unsequenced ‘*E. palauensis*’ from Palauan *T. swinhoei*^{23,24}, we compared the reported four 16S rRNA gene sequences from this sponge to our data. Only one (AF130847) exceeded 1,300 nt and was included in the phylogram (Supplementary Fig. 5b), which suggested a distinct phylotype. Further alignment including our shorter sequences (Supplementary Fig. 6a) and all four ‘*E. palauensis*’ 16S rRNA genes showed pairwise identities around 97% (Supplementary Fig. 6b), indicating no close relationship with any of our phylotypes despite the previous finding that the Palauan *T. swinhoei* contains NPs similar to those assigned to ‘*E. sarta*’ in *T. swinhoei* WA and WB^{23,24}.

Few shared gene clusters across BGC-rich symbionts

To evaluate the biosynthetic potential of the symbiont genomes, we searched for BGCs using antiSMASH⁵³ followed by manual reanalysis for validation and detection of orphan biosynthetic loci. Genomes contained a consistently high number of BGCs (fragments) that ranged from 10 to 42 (Fig. 2b and Supplementary Table 4). In fragmented

genome sequences, the BGC numbers for large, multimodular polyketide synthases (PKSs) or nonribosomal peptide synthetases (NRPSs) are overrepresented when BGCs are distributed over multiple contigs. To allow for a better comparison, Fig. 2b also shows catalytic domain counts including terminal domains for such multimodular assembly lines. For information about the chemical diversity across ‘Entotheonella’ variants, we assessed BGC similarities with the Biosynthetic Gene Similarity Clustering and Prospecting Engine (BiG-SCAPE)⁵⁴, which groups BGCs into gene cluster families (GCFs) and compares them to already characterized ones in the MIBiG database⁵⁵. The visualized data in Fig. 3, thus, allowed us to assign BGCs to putatively known or orphan compound types and compare the BGC diversity across phylotypes.

We detected a total of 493 BGCs or BGC fragments in the ‘Entotheonella’ and ‘Proxinella’ genomes, assigned to 369 GCFs and grouped by biosynthetic pathway classes: thiotemplate-based pathways (NRPSs and PKSs), ribosomally synthesized and post-translationally modified peptides (RiPPs), terpenes or other. Within this network, only seven links to previously characterized BGCs with MIBiG database entries were identified. All these belonged to BGCs already identified earlier in ‘Entotheonella’^{29,30,56} (Table 1).

Five additional GCFs matched previously assigned BGC types lacking MIBiG entries, namely konbamides (for example 20), keramamides (3), cyclotheonamides (6), nazumamide A (4) (Fig. 1 and Supplementary Fig. 1) and a partially characterized orphan proteusin from ‘*Ca. E. factor*’ TSY1 (ref. 29). The high percentage (96.7%) of unassigned GCFs and high GCF-to-BGC ratio (0.75) indicate considerable metabolic distinctness variation among the 18 newly analyzed symbionts.

In a previous study, we assigned the known polyketides and peptides from *T. swinhoei* Y to ‘*E. factor*’ TSY1 BGCs located on a plasmid (encoding pathways for onnamide and polytheonamide) and two genomic regions (encoding cyclotheonamide, konbamide, keramamide and nazumamide biosynthesis)²⁹. The co-occurring symbiont ‘*Ca. E. gemina*’ contained exclusively orphan BGCs. Unexpectedly, analysis of single-bacterial genomes of ‘*E. factor*’ TSYB1, ‘*E. gemina*’ TSYB2 and ‘*E. mitsugo*’ TSYB3 in *T. swinhoei* YB collected at the same Japanese location suggested that all known NP BGCs belong to ‘*E. gemina*’ TSYB2 instead of ‘*E. factor*’ TSYB1. Reanalyzing the data of the initial study regarding potential misassignments did not reveal errors. The data also contradicted a misassignment of 16S rRNA genes, as core genomes of each ‘*E. factor*’ and ‘*E. gemina*’ pair were largely identical, including the orphan BGCs. This suggests the BGCs and plasmid were either exchanged between ‘Entotheonella’ variants or differentially acquired from another source. Supporting this, BGC mobility was also observed in another study that detected ‘*E. factor*’ BGCs in ‘*E. sarta*’³¹. Beneficial properties of mobile BGCs, for example, in the context of host defense or colonization, might underlie the observed symbiont retention or switching patterns.

To further assess chemical diversity, we manually reanalyzed BiG-SCAPE results (Fig. 3a) and evaluated BGC similarities across phylotypes. All three ‘*E. sarta*’ variants (TCWA1, TCWB1 and TCBA2) from blue and white *Theonella* sponges share a substantial BGC repertoire, that is, 7–22 BGCs per symbiont pair (Supplementary Fig. 7). These include BGCs for related misakinolides and swinholides and a BGC for theonellamide. Genomes of different candidate species, however, contain mostly unique BGC sets (Supplementary Fig. 7).

Manual inspection and clinker⁵⁷ analysis of BGCs classified as shared by BiG-SCAPE revealed that many grouped thiotemplated enzymes show only partial similarities (T-GCF-02 to T-GCF-09; Supplementary Figs. 8–10), while some related pathways were not grouped by BiG-SCAPE. One example is a staphyloxanthin-like BGC discovered in an earlier ‘Entotheonella’ study (named theoxanthin BGC)³⁵. This BGC differs from typical NP clusters and, therefore, was missed in the antiSMASH analysis underlying the BiG-SCAPE network.

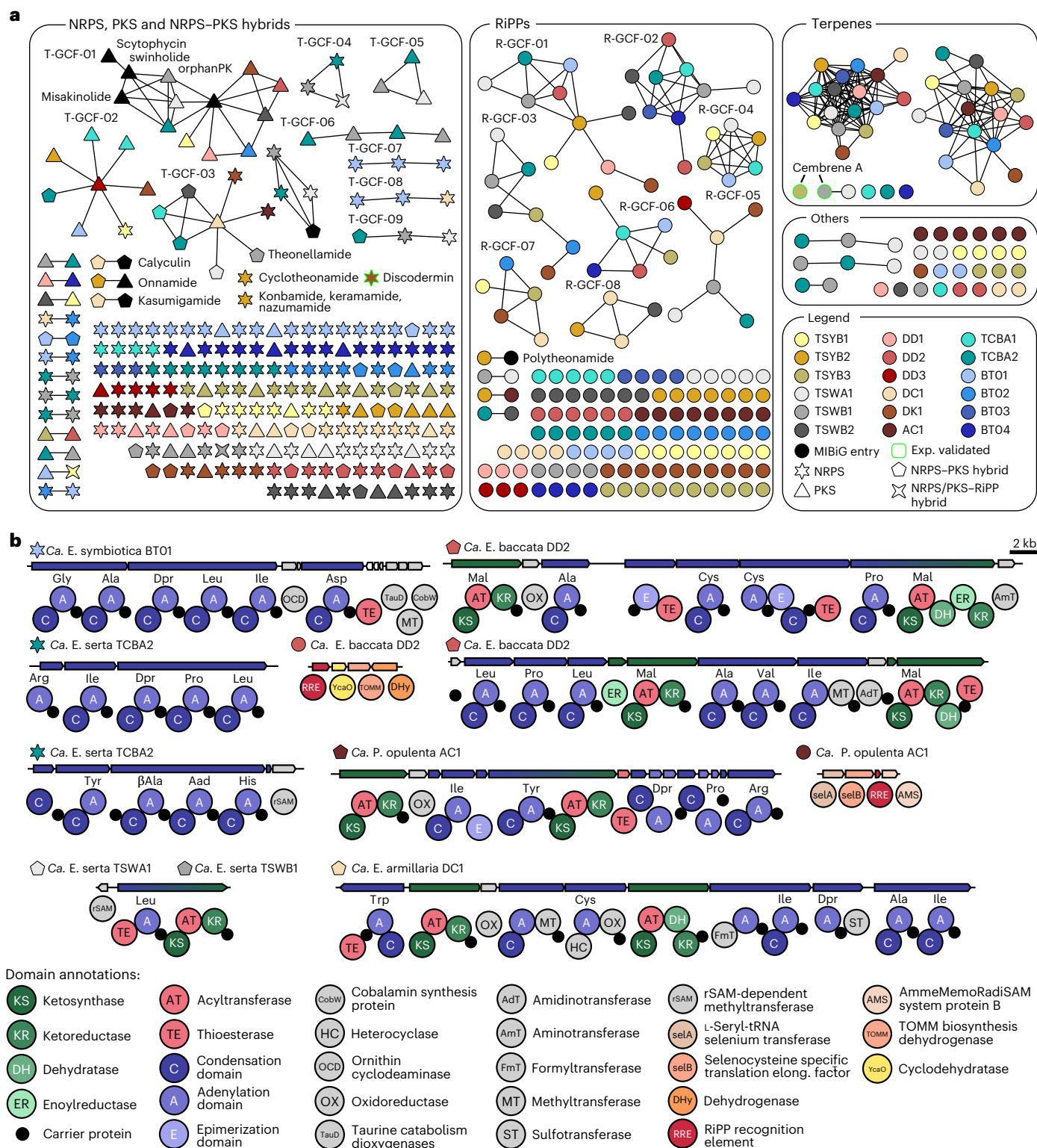


Fig. 3 | Biosynthetic potential analysis of all genomes. a, BiG-SCAPE output of a BGC network divided into the major NP classes: thiotemplate, RiPPs, terpenes and others. If two shapes are connected, their BGC (fragments) are similar to each other. The various colors and shapes refer to the genome and thiotemplate enzyme family, respectively, as described in the legend. **b**, Selected BGCs and

their encoded biosynthetic domains are shown. For acyltransferase (AT) and A domains, the substrate selectivity as predicted by antiSMASH is written above the domain. The protein family of domains is provided in the domain annotations. Dpr: 2,3-diaminopropionic acid.

We manually identified such BGCs in ten of 18 genomes and analyzed their relatedness using clinker⁵⁷ (Supplementary Fig. 11). This showed highly conserved architectures that might, given their prevalence,

indicate an important function for this candidate genus, possibly similar to staphyloxanthins that serve as antioxidant virulence factors during *Staphylococcus aureus* host colonization⁵⁸.

Among the remaining highly similar loci, several were classified as RiPP-type, RiPP-like or RRE-containing by antiSMASH⁵⁹ (Supplementary Figs. 12–15). However, all lacked identifiable precursor peptides, leaving their involvement in NP biosynthesis unclear. The few BGCs that were clearly shared by multiple phylotypes had architectures suggesting involvement in primary metabolic processes, that is, hopanoid, carotenoid, ectoine and pyrroloquinoline quinone biosynthesis. Additionally, a type III PKS BGC (Supplementary Fig. 16) found in nine of the 18 genomes was previously shown to encode biosynthesis of alkyl resorcinols and hydroquinones that might function as redox cofactors⁵⁶. Thus, many antiSMASH⁵³-detected shared BGC types likely belong to primary rather than secondary metabolism. An exception may be a BGC (T-GCF-03) encoding a bimodular hybrid NRPS-PKS with an unusual N-terminal thioesterase (TE) domain identified in eight genomes (Supplementary Fig. 9). Similarity searches revealed related enzymes with identical architecture in >50 phylogenetically diverse bacteria, mostly derived from eukaryotic hosts. However, these BGCs are uncharacterized.

We also checked the MOTUs database, the largest bacterial genome repository at the time of writing, for additional ‘Tectomicrobia’ members, finding 66 MAGs (Supplementary Table 5) beyond those reported here or in published work^{29,30,60}. Of these, 64 are primarily non-‘Entothaeonella’ members containing 3–5 shared BGCs that appear to belong to primary metabolites (carotenoid, hopanoid, ladderane and PKS-like type I fatty acid synthase) or single NRPS modules, indicating low chemical diversity. Of the two remaining MAGs, both assigned to ‘Entothaeonella’, one from a sponge metagenome contains four contigs with PKS or NRPS genes. The second MAG remained BGC rich after manual curation to eliminate BGCs from primary metabolism (12 NP contigs and 21 in total). This MAG originated from soil, supporting previous 16S rRNA data suggesting the existence of terrestrial BGC-rich members of this candidate genus³⁹.

In conclusion, the BGC analysis revealed high diversity and variability in predicted NP biosynthetic pathways and structures among the analyzed symbionts, with few BGCs assigned to known compounds. These findings warranted a closer examination of orphan pathways at the *in silico* and functional level.

BGC candidates for orphan compounds and sponge cytotoxins

In our dataset, 357 of the 369 GCFs lacked assigned NPs. Of these, 310 represented unique BGCs. Such singletons were found in every ‘Entothaeonella’ genome, suggesting, together with the numerous phylotypes encountered in this candidate genus, a large NP discovery resource. Examples of BGCs of as-yet unknown function are shown in Fig. 3b. They include several NRPS systems, one of them associated with a radical S-adenosylmethionine (rSAM) C-methyltransferase homolog as an unusual enzyme combination; another BGC with this feature is discussed below. Among the putative RiPP BGCs, an operon stood out that encodes a dioxygenase-RiPP recognition element fusion enzyme and homologs of the selenocysteine proteins SclA and SclB, suggesting noncanonical biochemistry. This BGC was also present in the ‘*Ca. P. opulenta*’ AC1 genome, along with BGCs for a keramamide-like NRPS, further RiPPs and other compound types.

With sequencing data from *D. dissoluta* available, we searched for a BGC candidate for discodermolide, an anticancer polyketide that had reached phase 1 clinical trials⁶¹. Analysis of methanolic sponge extracts confirmed the compound in the collected specimen (Supplementary Figs. 17 and 18). However, none of the three ‘Entothaeonella’ genomes of *D. dissoluta* contained a convincing candidate. As we only sequenced the enriched filamentous symbiont fraction, the discodermolide producer might be an organism distinct from ‘Entothaeonella’. Another missing BGC was the one encoding the biosynthetic pathway of discokiolides, depsipeptides reported from *D. kiiensis* collected at a different location from ours⁶². In agreement, an analysis of *D. kiiensis* extracts did not suggest that our specimens contained these compounds.

In contrast to the missing discodermolide genes, we found three additional BGCs that architecturally matched reported sponge NPs. *D. calyx* from Shikine-Jima contains cytotoxic calyxamides (for example, **14**; Fig. 1)⁶³, cyclic peptides with a formyl starter, a thiazole unit and two polyketide-like extensions, including an unusual C1 extension also found in keramamides (for example, **3**; Supplementary Fig. 1)⁶⁴. Correspondingly, the ‘*E. armillaria*’ DC1 genome contains two regions with a keramamide-type BGC matching the calyxamide structure (Fig. 3b, Supplementary Tables 6 and 7 and Extended Data Fig. 1), suggesting this symbiont as the source. In the ‘*E. monilis*’ DK1 genome from *D. kiiensis*, we identified BGCs matching the cytotoxic lipodiscamides (for example, **20**; Fig. 1) and the discodermidin antibiotics (**15–19**), known compounds from this sponge^{40,41,65–69}. As a lipodiscamide candidate, the *lpc* BGC in ‘*E. monilis*’ DK1 encodes a PKS–NRPS machinery that shows perfect architectural agreement with the polyketide-peptide hybrid structure of **20** (Fig. 4a, Supplementary Tables 7 and 8 and Extended Data Fig. 2). This includes a characteristic NRPS module with a ketoreductase (KR) domain, previously reported to generate α -hydroxyacid residues, as present in the hydroxyisovalerate ester moiety of **20** (ref. 70). Four PKS modules are predicted to catalyze four elongations of a methylactenoyl starter with methoxy and geminal dimethyl modifications introduced by *O*-methyltransferase and *C*-methyltransferase domains, respectively (Extended Data Fig. 2). The BGC also encodes a sulfotransferase homolog, consistent with the sulfonated lipodiscamides⁴¹. Another BGC (*dsc*) in ‘*E. monilis*’ DK1 encodes four NRPS proteins totaling 14 modules (Fig. 4b and Supplementary Table 9), including a predicted loading module with a formyltransferase domain. This feature and the overall order and predicted specificities (Fig. 4b and Supplementary Table 7) of adenylation (A) domains fit well with the tetradecapeptide structure of discodermidins (**15–19**; Figs. 1 and 4b and Extended Data Fig. 3). The only deviation of this prediction from the final chemical structure is aspartic acid (Asp) as a substrate for the A domain in module eight of DscC instead of cysteic acid (Cya) present in discodermidins. Additionally, epimerase and *N*-methyltransferase domains in some modules align with D-amino acids and *N*-methylated peptide bonds in discodermidins.

Collectively, these analyses suggest that most compounds previously reported from the sponges are produced by ‘Entothaeonella’, with the possible exception of discodermolides. To complement our *in silico* studies with functional data, we selected one tentatively assigned (*dsc*) and one orphan gene locus (*cmb*) for biochemical enzyme studies.

RiPP-like modification in nonribosomal biosynthesis

An unusual feature of discodermidins (**15–19**; Figs. 1 and 4b) is the presence of variants with optional *C*-methylations at three positions that contain alanine and valine in nonmethylated congeners. Because of these structural variants, we speculated that *C*-methylation might occur late in biosynthesis rather than through incorporation of premethylated amino acids. The *dsc* BGC encodes the protein DscE with homology to cobalamin-dependent rSAM methyltransferases, which typically catalyze radical *C*-methylations^{71,72}. An extreme example is 16–17 *C*-methylations in polytheonamides (**8**) from ‘*E. factor*’ (refs. 32,73–76), peptides superficially similar to discodermidins in their *t*-leucine residues and alternating DL-configurations. However, polytheonamides are RiPPs in contrast to the nonribosomally synthesized discodermidins.

To interrogate the function of DscE, we reisolated the nonmethylated congener discodermidin D (**18**) from *D. kiiensis* as a substrate for *in vitro* methylation. DscE was prepared by aerobically expressing its codon-optimized gene in *Escherichia coli* Tuner (DE3) as an N-terminally His₆-tagged variant. The gene was coexpressed with the *Azotobacter vinelandii* *isc* operon for iron–sulfur cluster biosynthesis⁷⁷ and native *btuCEDFB* genes for cobalamin uptake, previously reported to aid the production of B₁₂-dependent rSAM enzymes⁷⁸.

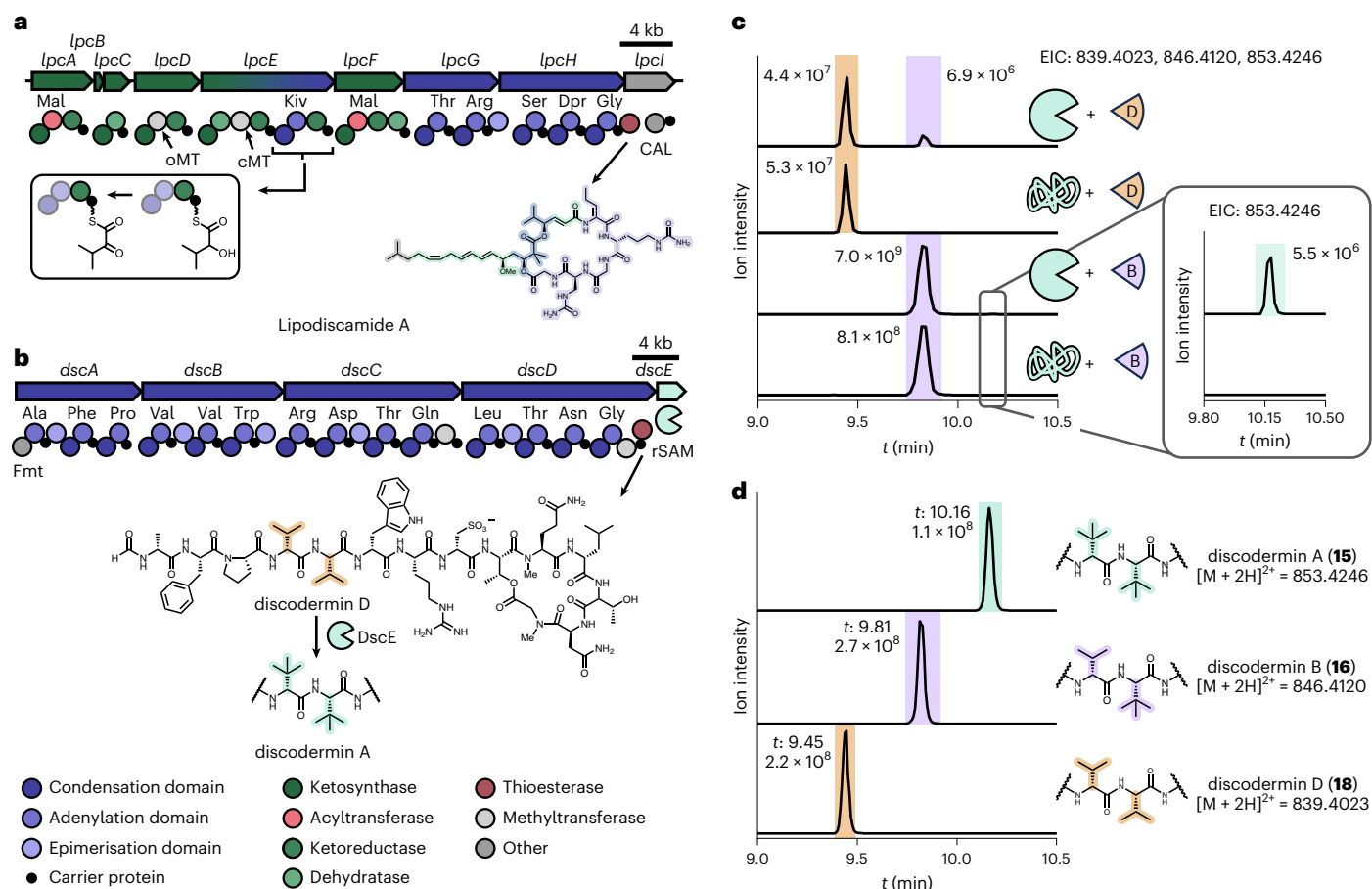


Fig. 4 | In silico prediction of lipodiscamide biosynthesis and biochemical study of discodermin biosynthesis. **a**, Putative lipodiscamide BGC. Shown above the NRPS and PKS proteins are the predicted substrate specificities for the respective A and AT domains. The module architectures and predicted substrate specificities fit well to the chemical structure of lipodiscamide A. As an example, the KR-catalyzed reaction of α -ketoisovaleric acid to the corresponding α -hydroxyl moiety is shown. A full biosynthetic scheme is provided in Extended Data Fig. 2. **b**, BGC encoding the discodermin NRPS. Shown above the NRPS proteins are the predicted substrate specificities for the respective A domains. A full biosynthetic scheme is provided in Extended Data Fig. 3. The reaction of the

rSAM methyltransferase DscE acting on discodermin intermediates is shown. **c**, HPLC-HRMS traces of in vitro reconstitutions with native (pacman shape) or heat-denatured DscE (noodle shape) and discodermin D or discodermin B (wedges). Depicted are extracted ion chromatograms (EICs). Total ion counts of HPLC-HRMS runs are provided in Supplementary Fig. 21. **d**, HPLC-HRMS traces of isolated discodermin A (blue), B (purple) and D (orange). Serving as analytical standards for reactions shown in **c** (color-coded similarly). HPLC-HRMS traces on all further control reactions are provided in Supplementary Fig. 23. Kiv, α -ketoisovaleric acid; CAL, CoA-acyl ligase; Fmt, formyltransferase.

After anaerobic purification using nickel affinity chromatography (Supplementary Fig. 19), iron-sulfur clusters and the cobalamin cofactor were anaerobically reconstituted by adding iron and sulfur sources (ammonium iron(II) sulfate, L-cysteine, cysteine desulfurase (IscS) and pyridoxal phosphate) and methylcobalamin (MeCbl) (Supplementary Fig. 20).

We tested the activity of DscE by incubation with unmethylated discodermin D, SAM, MeCbl and a reductant system (methyl viologen, DTT and NADPH) under anaerobic conditions. High-performance liquid chromatography-high-resolution mass spectrometry (HPLC-HRMS) analysis showed formation of a product with a 14-Da mass increase (Fig. 4c and Supplementary Fig. 21), localized to V5 by MS² fragmentation (Extended Data Fig. 4). Comparison to authentic standards from *D. kiliensis* confirmed its identity as discodermin B (16) (Fig. 4d and Supplementary Figs. 22 and 23). When the methylation assay was repeated with monomethylated discodermin B (16) from *D. kiliensis* as substrate, small amounts of dimethylated peptide were detected with properties identical to discodermin A (15) (Fig. 4c).

These data on the RiPP-like discodermin modification and the good correspondence between NRPS architecture and discodermin structure support biosynthesis of these peptides by ‘*E. monilis*’ DK1.

The origin of *N*-ethylglycine (EtGly) and Cya building blocks remains unclear. EtGly formation by C-methylation of Al was not observed in our assays. EtGly, a known transamination product of the central metabolite 2-ketobutyrate, might be directly incorporated by the NRPS. Similarly, Cya, for which biosynthetic gene candidates were not identified, might first form as a free amino acid. This is supported by its structural similarity to Asp, the predicted A domain substrate of the corresponding NRPS module. While no gene candidates for previously described Cya biosynthetic enzymes^{79,80} were detected in the ‘*E. monilis*’ genome, Cya might be produced by another member of the sponge holobiont, including the host, as this amino acid is a known sponge metabolite⁸¹. *tert*-Leucine is a residue of several other predicted NRPS/PKS products and might likewise be generated by radical C-methylation, notably the promising sponge-derived anticancer drug candidates plocabulin^{82,83} and hemiasterlin⁸⁴, both with unknown biosynthetic origin.

Characterization of an orphan terpene pathway

A range of terpene NPs have been reported from *T. swinhoei*, including the isonitrile diterpene amitorine A⁸⁵ and several steroids⁸⁶. Inspection of our genome data revealed 50 predicted terpene biosynthetic loci across all 18 genomes (Fig. 2b). However, most exhibited architectures

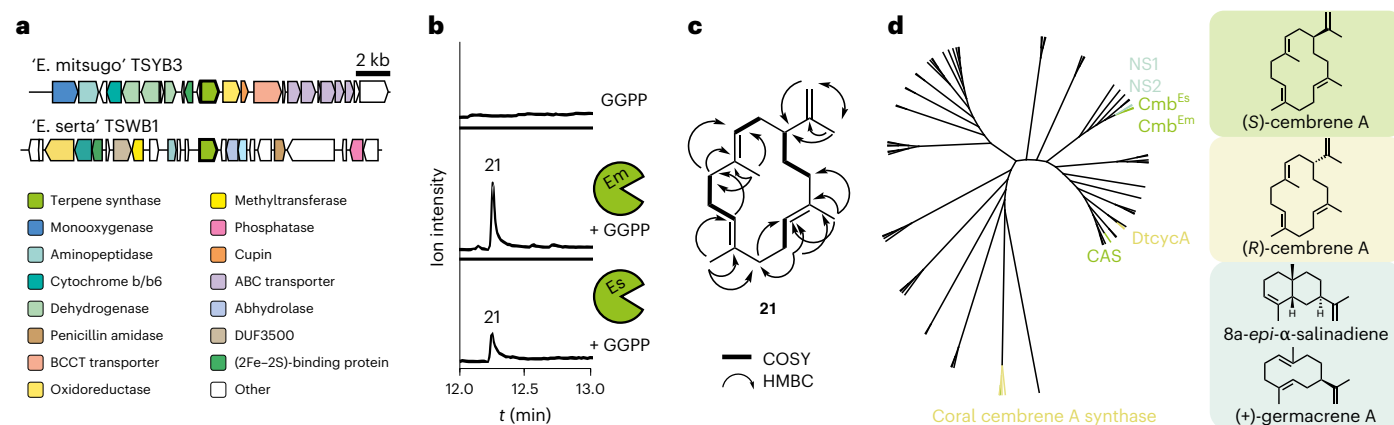


Fig. 5 | Terpene synthase responsible for cembrene A biosynthesis. a, BGCs containing the *cmb* gene (terpene synthase) found in the genomes of 'E. mitsugo' TSYB3 and 'E. sarta' TSWB1. **b**, GC-MS traces of in vitro reconstitutions of Cmb with GGPP. **c**, Structure elucidation of **21**, product formed in the reaction of Cmb with GGPP. **d**, Phylogenetic tree of selected bacterial terpene synthases and the homolog catalyzing cembrene A (**21**) formation in soft corals^{94,95}.

GenBank identifiers of the sequence used to construct this tree are shown in Supplementary Table 11. NS1, 8a-epi-α-selinene synthase from *Nostoc* sp. PCC 7120; NS2, (+)-germacrene A synthase from *Nostoc punctiforme* PCC 73102; DtcycA, (R)-cembrene synthase from *Streptomyces* sp. SANK 60404; CAS, (S)-cembrene synthase from *Allokutzneria albatra*.

that suggest carotenoid or hopanoid biosynthesis. This agrees with our previous detection of carotenoids in single 'Entotheonella' filaments using Raman microscopy⁴⁶.

Among terpene loci lacking carotenoid-type and hopanoid-type genes, we identified genome regions in the *Theonella* sp. 1 BA symbiont 'E. sarta' TCBA2, the *T. swinhoei* YB symbiont 'E. mitsugo' TSYB3, the *T. swinhoei* WB symbiont 'E. sarta' TSWB1 and the *T. swinhoei* WA symbiont 'E. sarta' TSWA1 with closely related genes encoding a predicted class I terpene synthase, termed Cmb. Each *cmb* gene was embedded in a distinct genomic environment with unclear roles in NP biosynthesis (Fig. 5a and Supplementary Fig. 24). To assess the terpene synthase function, we selected two of the four enzymes for further analyses. We heterologously expressed codon-optimized genes from 'E. mitsugo' TSYB3 (named *cmb*^{Em}) and 'E. sarta' TSWB1 (*cmb*^{Es}) (Supplementary Table 10) in *E. coli* BL21 (DE3) as N-terminally His₆-tagged proteins (Supplementary Fig. 25a). Incubation with geranyl pyrophosphate (GPP), farnesyl pyrophosphate (FPP) or geranyl-GPP (GGPP) in vitro and analysis by gas chromatography (GC)-MS (Supplementary Fig. 25b) revealed the formation of geraniol (Supplementary Fig. 26) and several sesquiterpenes (Supplementary Fig. 27) when incubated with GPP and FPP, respectively.

However, incubating GGPP with either Cmb^{Em} or Cmb^{Es} resulted in the formation of a single compound with identical retention times and mass spectra in both cases (Fig. 5b and Extended Data Fig. 5). This compound was purified from a preparative-scale enzymatic reaction using Cmb^{Es} from 'E. sarta', yielding 0.8 mg of product. Nuclear magnetic resonance (NMR)-based structure elucidation (Fig. 5c, Extended Data Table 1 and Supplementary Figs. 28–32) identified the terpene as cembrene A (**21**). Optical rotation measurements and comparison to literature values^{87,88} established the compound as (S)-cembrene A ([α]_D²⁰ = +5.7).

Discussion

This study provides deeper insights into sponge symbionts of the candidate genus 'Entotheonella'. We show that large and variable sets of unique BGCs are a consistent feature across the investigated members of this lineage. Moreover, among four types of biosynthetically unassigned polyketides and modified peptides previously reported from the selected sponges (discodermolides, calyxamides, lipodiscamides and discodermins), three were bioinformatically or functionally linked to 'Entotheonella' BGCs, supporting widespread roles of these symbionts as producers of structurally complex sponge NPs. The data also

suggest the existence of a second BGC-rich taxon, 'Proxinella', within the candidate phylum 'Tectomicrobia' (ref. 39). While defensive roles were demonstrated for some sponge NPs⁸, rigorous ecological studies are needed to test this hypothesis for the 'Entotheonella' compounds. Alternative functions may be to mediate interactions within the microbiome or to aid the producer in sponge colonization, for example.

Various non-'Entotheonella' symbionts were previously bioinformatically or functionally linked to sponge NPs. These include a Chloroflexi member in *T. swinhoei* as the aurantioside source⁴⁶, various polyketide-producing bacteria in *Mycale hentscheli*^{20,21}, cyanobacteria producing halogenated compounds in dysideid sponges⁸⁹, an intracellular renieramyin-producing gammaproteobacterium in a *Haliclona* sponge⁹⁰ and diverse bacteria producing halogenated RiPPs⁹¹. In addition, sponge hosts have been demonstrated to synthesize some terpenes¹² and peptides^{13,92}. Our identification of the 'Entotheonella' product cembrene A was unexpected. Cembranoids are known from various organisms including corals and plants⁹³. Recent biochemical studies assigned cembrene biosynthesis in sponges and octocorals^{12,94,95} to the animals. Our data show that terpenes of marine invertebrates can have diverse origins even for identical compounds. In agreement, cembrene A cyclases were also reported from actinomycetes^{96,97}. However, their sequences greatly differ, suggesting convergent evolution (Fig. 5d and Supplementary Fig. 33). Phylogenetically, the two 'Entotheonella' homologs were more similar to cyanobacterial 8a-epi-α-selinene and germacrene A cyclase than to other cembrene cyclases (Fig. 5d).

To date, 'Entotheonella' cultivation attempts have been unsuccessful except for one report on a mixed culture²³. The genome data provided here and in earlier studies^{29,35} might aid targeted cultivation approaches to access the diverse chemistry of this talented producer taxon. The available genetic information on assigned and orphan pathways might also enable additional supply strategies, including heterologous BGC expression and the targeted search for alternative culturable producer organisms containing homologous genes⁹⁸, approaches likely to become increasingly successful with current and future genome initiatives.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41589-025-02066-0>.

References

- Carroll, A. R., Copp, B. R., Grkovic, T., Keyzers, R. A. & Prinsep, M. R. Marine natural products. *Nat. Prod. Rep.* **41**, 162–207 (2024).
- Paul, V. J., Freeman, C. J. & Agarwal, V. Chemical ecology of marine sponges: new opportunities through ‘-omics’. *Integr. Comp. Biol.* **59**, 765–776 (2019).
- Slaby, B. M., Hackl, T., Horn, H., Bayer, K. & Hentschel, U. Metagenomic binning of a marine sponge microbiome reveals unity in defense but metabolic specialization. *ISME J.* **11**, 2465–2478 (2017).
- Perdicaris, S., Vlachogianni, T. & Valavanidis, A. Bioactive natural substances from marine sponges: new developments and prospects for future pharmaceuticals. *Nat. Prod. Chem. Res.* **1**, 1000115 (2013).
- Becerro, M. A., Thacker, R. W., Turon, X., Uriz, M. J. & Paul, V. J. Biogeography of sponge chemical ecology: comparisons of tropical and temperate defenses. *Oecologia* **135**, 91–101 (2003).
- Engel, S., Jensen, P. R. & Fenical, W. Chemical ecology of marine microbial defense. *J. Chem. Ecol.* **28**, 1971–1985 (2002).
- Hamoda, A. M. et al. Evolutionary relevance of metabolite production in relation to marine sponge bacteria symbiont. *Appl. Microbiol. Biotechnol.* **107**, 5225–5240 (2023).
- Pawlik, J. R. The chemical ecology of sponges on Caribbean reefs: natural products shape natural systems. *BioScience* **61**, 888–898 (2011).
- Fazeela Mahaboob Begum, S. M. & Hemalatha, S. Marine natural products—a vital source of novel biotherapeutics. *Curr. Pharmacol. Rep.* **8**, 339–349 (2022).
- Proksch, P., Edrada-Ebel, R. A. & Ebel, R. Drugs from the sea—opportunities and obstacles. *Mar. Drugs* **1**, 5–17 (2003).
- Okamura, Y., et al. Screening of neutrophil activating factors from a metagenome library of sponge-associated bacteria. *Mar. Drugs* **19**, 427 (2021).
- Wilson, K., et al. Terpene biosynthesis in marine sponge animals. *Proc. Natl Acad. Sci. USA* **120**, e2220934120 (2023).
- Lin, Z., Agarwal, V., Cong, Y., Pomponi, S. A. & Schmidt, E. W. Short macrocyclic peptides in sponge genomes. *Proc. Natl Acad. Sci. USA* **121**, e2314383121 (2024).
- Piel, J. Metabolites from symbiotic bacteria. *Nat. Prod. Rep.* **21**, 519–538 (2004).
- Hentschel, U., Piel, J., Degnan, S. M. & Taylor, M. W. Genomic insights into the marine sponge microbiome. *Nat. Rev. Microbiol.* **10**, 641–654 (2012).
- Maslin, M., Gaertner-Mazouni, N., Debitus, C., Joy, N. & Ho, R. Marine sponge aquaculture towards drug development: an ongoing history of technical, ecological, chemical considerations and challenges. *Aquac. Rep.* **21**, 100813 (2021).
- Galitz, A., Nakao, Y., Schupp, P. J., Wörheide, G. & Erpenbeck, D. A soft spot for chemistry—current taxonomic and evolutionary implications of sponge secondary metabolite distribution. *Mar. Drugs* **19**, 448 (2021).
- de Oliveira, B. F. R., Carr, C. M., Dobson, A. D. W. & Laport, M. S. Harnessing the sponge microbiome for industrial biocatalysts. *Appl. Microbiol. Biotechnol.* **104**, 8131–8154 (2020).
- Cichewicz, R. H., Valeriote, F. A. & Crews, P. Psymberin, a potent sponge-derived cytotoxin from *Psammocinia* distantly related to the pederin family. *Org. Lett.* **6**, 1951–1954 (2004).
- Rust, M. et al. A multiproducer microbiome generates chemical diversity in the marine sponge *Mycale hentscheli*. *Proc. Natl Acad. Sci. USA* **117**, 9508–9518 (2020).
- Storey, M. A. et al. Metagenomic exploration of the marine sponge *Mycale hentscheli* uncovers multiple polyketide-producing bacterial symbionts. *mBio* **11**, e02997-19 (2020).
- Nakao, Y. et al. Identification of renieramycin A as an antileishmanial substance in a marine sponge *Neopetrosia* sp. *Mar. Drugs* **2**, 55–62 (2004).
- Schmidt, E. W., Obraztsova, A. Y., Davidson, S. K., Faulkner, D. J. & Haygood, M. G. 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).
- Bewley, C. A. & Faulkner, D. J. Lithistid sponges: star performers or hosts to the stars. *Angew. Chem. Int. Ed. Engl.* **37**, 2162–2178 (1998).
- Bewley, C. A., Holland, N. D. & Faulkner, D. J. Two classes of metabolites from *Theonella swinhoei* are localized in distinct populations of bacterial symbionts. *Experientia* **52**, 716–722 (1996).
- Schmidt, E. W., Bewley, C. A. & Faulkner, D. J. Theopalauamide, a bicyclic glycopeptide from filamentous bacterial symbionts of the lithistid sponge *Theonella swinhoei* from Palau and Mozambique. *J. Org. Chem.* **63**, 1254–1258 (1998).
- Fusetani, N. & Matsunaga, S. Bioactive sponge peptides. *Chem. Rev.* **93**, 1793–1806 (1993).
- D’Auria, M. V., Zampella, A. & Zollo, F. The chemistry of lithistid sponge: a spectacular source of new metabolites. *Stud. Nat. Prod. Chem.* **26**, 1175–1258 (2002).
- Wilson, M. C. et al. An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* **506**, 58–62 (2014).
- Mori, T. et al. Single-bacterial genomics validates rich and varied specialized metabolism of uncultivated *Entotheonella* sponge symbionts. *Proc. Natl Acad. Sci. USA* **115**, 1718–1723 (2018).
- Yamabe, S. et al. Metagenomic insights reveal unrecognized diversity of *Entotheonella* in Japanese *Theonella* sponges. *Mar. Biotechnol.* **26**, 1009–1016 (2024).
- Freeman, M. F. et al. Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides. *Science* **338**, 387–390 (2012).
- Piel, J. et al. Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proc. Natl Acad. Sci. USA* **101**, 16222–16227 (2004).
- Ueoka, R. et al. Metabolic and evolutionary origin of actin-binding polyketides from diverse organisms. *Nat. Chem. Biol.* **11**, 705–712 (2015).
- Lackner, G., Peters, E. E., Helfrich, E. J. N. & Piel, J. Insights into the lifestyle of uncultured bacterial natural product factories associated with marine sponges. *Proc. Natl Acad. Sci. USA* **114**, E347–E356 (2017).
- Wakimoto, T. et al. Calyculin biogenesis from a pyrophosphate protoxin produced by a sponge symbiont. *Nat. Chem. Biol.* **10**, 648–655 (2014).
- Nakashima, Y., Egami, Y., Kimura, M., Wakimoto, T. & Abe, I. Metagenomic analysis of the sponge *Discodermia* reveals the production of the cyanobacterial natural product kasumigamide by ‘*Entotheonella*’. *PLoS ONE* **11**, e0164468 (2016).
- Fisch, K. M. et al. Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. *Nat. Chem. Biol.* **5**, 494–501 (2009).
- Peters, E. E., et al. Distribution and diversity of ‘Tectomicrobia’, a deep-branching uncultivated bacterial lineage harboring rich producers of bioactive metabolites. *ISME Commun.* **3**, 50 (2023).
- Tan, K. C., Wakimoto, T. & Abe, I. Lipodiscamides A–C, new cytotoxic lipopeptides from *Discodermia kiiensis*. *Org. Lett.* **16**, 3256–3259 (2014).
- Tan, K. C., Wakimoto, T. & Abe, I. Sulfoureido lipopeptides from the marine sponge *Discodermia kiiensis*. *J. Nat. Prod.* **79**, 2418–2422 (2016).
- Kato, Y. et al. Calyculin A, a novel antitumor metabolite from the marine sponge *Discodermia calyx*. *J. Am. Chem. Soc.* **108**, 2780–2781 (1986).

43. Brück, W. M., Sennett, S. H., Pomponi, S. A., Willenz, P. & McCarthy, P. J. Identification of the bacterial symbiont *Entotheonella* sp. in the mesophyl of the marine sponge *Discodermia* sp. *ISME J.* **2**, 335–339 (2008).
44. Gunasekera, S. P., Gunasekera, M., Longley, R. E. & Schulte, G. K. Discodermolide: a new bioactive polyhydroxylated lactone from the marine sponge *Discodermia dissoluta*. *J. Org. Chem.* **55**, 4912–4915 (1990).
45. Gunasekera, S. P., Paul, G. K., Longley, R. E., Isbrucker, R. A. & Pomponi, S. A. Five new discodermolide analogues from the marine sponge *Discodermia* species. *J. Nat. Prod.* **65**, 1643–1648 (2002).
46. Kogawa, M., et al. Single-cell metabolite detection and genomics reveals uncultivated talented producer. *PNAS Nexus* **1**, pgab007 (2022).
47. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
48. Robbins, S. J. et al. A genomic view of the microbiome of coral reef demosponges. *ISME J.* **15**, 1641–1654 (2021).
49. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).
50. Alanjary, M., Steinke, K. & Ziemert, N. AutoMLST: an automated web server for generating multi-locus species trees highlighting natural product potential. *Nucleic Acids Res.* **47**, W276–W282 (2019).
51. Parks, D. H. et al. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* **50**, D785–D794 (2022).
52. Kim, H., Ahn, J., Kim, J. & Kang, H. S. Metagenomic insights and biosynthetic potential of *Candidatus* *Entotheonella* symbiont associated with *Halichondria* marine sponges. *Microbiol. Spectr.* **13**, e02355-24 (2025).
53. Blin, K. et al. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **49**, W29–W35 (2021).
54. Navarro-Muñoz, J. C. et al. A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* **16**, 60–68 (2020).
55. Terlouw, B. R. et al. MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Res.* **51**, D603–D610 (2023).
56. Reiter, S., Cahn, J. K. B., Wiebach, V., Ueoka, R. & Piel, J. Characterization of an orphan type III polyketide synthase conserved in uncultivated ‘*Entotheonella*’ sponge symbionts. *ChemBioChem* **21**, 564–571 (2020).
57. Gilchrist, C. L. M. & Chooi, Y. H. clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* **37**, 2473–2475 (2021).
58. Clauditz, A., Resch, A., Wieland, K. P., Peschel, A. & Götz, F. Staphyloxanthin plays a role in the fitness of *Staphylococcus aureus* and its ability to cope with oxidative stress. *Infect. Immun.* **74**, 4950–4953 (2006).
59. Blin, K. et al. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.* **51**, W46–W50 (2023).
60. Liu, F., Li, J., Feng, G. & Li, Z. New genomic insights into ‘*Entotheonella*’ symbionts in *Theonella swinhoei*: mixotrophy, anaerobic adaptation, resilience, and interaction. *Front. Microbiol.* **7**, 1333 (2016).
61. Mita, A. et al. A phase I pharmacokinetic (PK) trial of XAA296A (discodermolide) administered every 3 weeks to adult patients with advanced solid malignancies. *J. Clin. Oncol.* **22**, 2025–2025 (2004).
62. Tada, H., Tozjo, T., Terui, Y. & Fumiaki, H. Discokiolides. Cytotoxic cyclic depsipeptides from the marine sponge *Discodermia kiiensis*. *Chem. Lett.* **21**, 431–434 (1992).
63. Kimura, M. et al. Calyxamides A and B, cytotoxic cyclic peptides from the marine sponge *Discodermia calyx*. *J. Nat. Prod.* **75**, 290–294 (2012).
64. Smith, D. R. M. et al. An unusual flavin-dependent halogenase from the metagenome of the marine sponge *Theonella swinhoei* WA. *ACS Chem. Biol.* **12**, 1281–1287 (2017).
65. Matsunaga, S., Fusetani, N. & Konosu, S. Bioactive marine metabolites IV. Isolation and the amino acid composition of discodermin A, an antimicrobial peptide, from the marine sponge *Discodermia kiiensis*. *J. Nat. Prod.* **48**, 236–241 (1985).
66. Matsunaga, S., Fusetani, N. & Konosu, S. Bioactive marine metabolites VI. Structure elucidation of discodermin A, an antimicrobial peptide from the marine sponge *Discodermia kiiensis*. *Tetrahedron Lett.* **25**, 5165–5168 (1984).
67. Matsunaga, S., Fusetani, N. & Konosu, S. Bioactive marine metabolites VII. Structures of discodermins B, C and D, antimicrobial peptides from the marine sponge *Discodermia kiiensis*. *Tetrahedron Lett.* **26**, 855–856 (1985).
68. Ryu, G., Matsunaga, S. & Fusetani, N. Discodermins F–H, cytotoxic and antimicrobial tetradecapeptides from the marine sponge *Discodermia kiiensis*: structure revision of discodermins A–D. *Tetrahedron* **50**, 13409–13416 (1994).
69. Ryu, G., Matsunaga, S. & Fusetani, N. Discodermin E, a cytotoxic and antimicrobial tetradecapeptide, from the marine sponge *Discodermia kiiensis*. *Tetrahedron Lett.* **35**, 8251–8254 (1994).
70. Magarvey, N. A., Ehling-Schulz, M. & Walsh, C. T. Characterization of the cereulide NRPS α -hydroxy acid specifying modules: activation of α -keto acids and chiral reduction on the assembly line. *J. Am. Chem. Soc.* **128**, 10698–10699 (2006).
71. Millera, D. V. et al. Radical S-adenosylmethionine methylases. in *Comprehensive Natural Products III* (eds Liu, H. W. & Begley, T. P.) (Elsevier, 2020).
72. Benjdia, A. & Berteau, O. B₁₂-dependent radical SAM enzymes: ever expanding structural and mechanistic diversity. *Curr. Opin. Struct. Biol.* **83**, 102725 (2023).
73. Bhushan, A., Egli, P. J., Peters, E. E., Freeman, M. F. & Piel, J. Genome mining- and synthetic biology-enabled production of hypermodified peptides. *Nat. Chem.* **11**, 931–939 (2019).
74. Hamada, T., Matsunaga, S., Yano, G. & Fusetani, N. Polytheonamides A and B, highly cytotoxic, linear polypeptides with unprecedented structural features, from the marine sponge, *Theonella swinhoei*. *J. Am. Chem. Soc.* **127**, 110–118 (2005).
75. Hamada, T., Sugawara, T., Matsunaga, S. & Fusetani, N. Polytheonamides, unprecedented highly cytotoxic polypeptides, from the marine sponge *Theonella swinhoei*: 1. Isolation and component amino acids. *Tetrahedron Lett.* **35**, 719–720 (1994).
76. Hamada, T., Sugawara, T., Matsunaga, S. & Fusetani, N. Polytheonamides, unprecedented highly cytotoxic polypeptides from the marine sponge *Theonella swinhoei*: 2. Structure elucidation. *Tetrahedron Lett.* **35**, 609–612 (1994).
77. Lanz, N. D. et al. RlmN and AtrB as models for the overproduction and characterization of radical SAM proteins. *Methods Enzymol.* **516**, 125–152 (2012).
78. Lanz, N. D. et al. Enhanced solubilization of class B radical S-adenosylmethionine methylases by improved cobalamin uptake in *Escherichia coli*. *Biochemistry* **57**, 1475–1490 (2018).
79. Takeda, K. et al. N-phenylacetylation and nonribosomal peptide synthetases with substrate promiscuity for biosynthesis of heptapeptide variants, JBIR-78 and JBIR-95. *ACS Chem. Biol.* **12**, 1813–1819 (2017).

80. Wang, M., Chen, D., Zhao, Q. & Liu, W. Isolation, structure elucidation, and biosynthesis of a cysteine-containing nonribosomal peptide in *Streptomyces lincolnensis*. *J. Org. Chem.* **83**, 7102–7108 (2018).
81. Hastings, J. et al. ChEBI in 2016: improved services and an expanding collection of metabolites. *Nucleic Acids Res.* **44**, D1214–D1219 (2016).
82. Costales-Carrera, A., et al. Plocabulin displays strong cytotoxic activity in a personalized colon cancer patient-derived 3D organoid assay. *Mar. Drugs* **17**, 648 (2019).
83. Martín, M. J. et al. Isolation and first total synthesis of PM050489 and PM060184, two new marine anticancer compounds. *J. Am. Chem. Soc.* **135**, 10164–10171 (2013).
84. Anderson, H. J., Coleman, J. E., Andersen, R. J. & Roberge, M. Cytotoxic peptides hemiasterlin, hemiasterlin A and hemiasterlin B induce mitotic arrest and abnormal spindle formation. *Cancer Chemother. Pharmacol.* **39**, 223–226 (1996).
85. Ota, K. et al. Amitorines A and B, nitrogenous diterpene metabolites of *Theonella swinhoei*: isolation, structure elucidation, and asymmetric synthesis. *J. Nat. Prod.* **79**, 996–1004 (2016).
86. Festa, C., De Marino, S., Zampella, A. & Fiorucci, S. *Theonella*: a treasure trove of structurally unique and biologically active sterols. *Mar. Drugs* **21**, 291 (2023).
87. Schwabe, R., Farkas, I. & Pfander, H. Synthese von (–)-(R)-nephthenol und (–)-(R)-cembren A. *Helv. Chim. Acta* **71**, 292–297 (1988).
88. Kato, T., Suzuki, M., Kobayashi, T. & Moore, B. P. Synthesis and pheromone activities of optically active neocembrenes and their geometrical isomers, (E,Z,E)- and (E,E,Z)-neocembrenes. *J. Org. Chem.* **45**, 1126–1130 (1980).
89. Agarwal, V. et al. Metagenomic discovery of polybrominated diphenyl ether biosynthesis by marine sponges. *Nat. Chem. Biol.* **13**, 537–543 (2017).
90. Tianero, M. D., Balaich, J. N. & Donia, M. S. Localized production of defence chemicals by intracellular symbionts of *Haliclona* sponges. *Nat. Microbiol.* **4**, 1149–1159 (2019).
91. Nguyen, N. A. et al. An obligate peptidyl brominase underlies the discovery of highly distributed biosynthetic gene clusters in marine sponge microbiomes. *J. Am. Chem. Soc.* **143**, 10221–10231 (2021).
92. Steffen, K. et al. Whole genome sequence of the deep-sea sponge *Geodia barretti* (Metazoa, Porifera, Demospongiae). *G3* **13**, jkad192 (2023).
93. Dauben, W. G., Thiessen, W. E. & Resnick, P. R. Cembrene, a 14-membered ring diterpene hydrocarbon. *J. Am. Chem. Soc.* **84**, 2015–2016 (1962).
94. Burkhardt, I., de Rond, T., Chen, P. Y. & Moore, B. S. Ancient plant-like terpene biosynthesis in corals. *Nat. Chem. Biol.* **18**, 664–669 (2022).
95. Scesa, P. D., Lin, Z. & Schmidt, E. W. Ancient defensive terpene biosynthetic gene clusters in the soft corals. *Nat. Chem. Biol.* **18**, 659–663 (2022).
96. Rinkel, J., Lauterbach, L., Rabe, P. & Dickschat, J. S. Two diterpene synthases for spiroalbatene and cembrene A from *Allokutzneria albata*. *Angew. Chem. Int. Ed. Engl.* **57**, 3238–3241 (2018).
97. Meguro, A., Tomita, T., Nishiyama, M. & Kuzuyama, T. Identification and characterization of bacterial diterpene cyclases that synthesize the cembrane skeleton. *ChemBioChem* **14**, 316–321 (2013).
98. Leopold-Messer, S. et al. Animal-associated marine Acidobacteria with a rich natural-product repertoire. *Chem* **9**, 3696–3713 (2023).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

¹Institute of Microbiology, Eidgenössische Technische Hochschule Zürich (ETH), Zurich, Switzerland. ²Department of Life Science and Medical Bioscience, Waseda University, Tokyo, Japan. ³Computational Bio Big-Data Open Innovative Laboratory, AIST-Waseda University, Tokyo, Japan. ⁴Research Organization for Nano & Life Innovation, Waseda University, Tokyo, Japan. ⁵Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan. ⁶Collaborative Research Institute for Innovative Microbiology, The University of Tokyo, Tokyo, Japan. ⁷Laboratory of Microbiology, Wageningen University and Research, Wageningen, The Netherlands. ⁸Laboratory of Natural Product Chemistry, Faculty of Pharmaceutical Sciences, Hokkaido University, Sapporo, Japan. ⁹Graduate School of Pharmaceutical Sciences, The University of Tokyo, Tokyo, Japan. ¹⁰Center for Biotechnology (CeBiTec), Bielefeld University, Bielefeld, Germany. ¹¹Bioinformatics Group, Wageningen University and Research, Wageningen, The Netherlands. ¹²Pharmacognosy, Department of Pharmaceutical Biosciences, BioMedical Center, Uppsala University, Uppsala, Sweden. ¹³Museum of Evolution, Uppsala University, Uppsala, Sweden. ¹⁴Bioprocess Engineering, Wageningen University and Research, Wageningen, The Netherlands. ¹⁵Harbor Branch Oceanographic Institute, Florida Atlantic University, Fort Pierce, FL, USA. ¹⁶School of Marine Biosciences, Kitasato University, Sagami-hara, Kanagawa, Japan. ¹⁷These authors contributed equally: Maria Dell, Masato Kogawa, Alena B. Streiff. ¹⁸These authors jointly supervised this work: Toshiyuki Wakimoto, Haruko Takeyama, Jörn Piel. ✉ e-mail: wakimoto@pharm.hokudai.ac.jp; haruko-takeyama@waseda.jp; jpiel@ethz.ch

Methods

General

Our research complied with all relevant ethical regulations.

The sample size for sponge samples was chosen on the basis of their availability: one for *A. cribrophora*, *D. calyx*, *D. kiiensis* and *T. swinhoei* WA; two for *T. swinhoei* WB and *Theonella* sp. 1 BA; three for *D. dissoluta* and *T. swinhoei* YB; four for *Theonella* sp. 2 BT. No statistical analyses were included in this study and none of the sponge specimens were excluded from our analyses.

Sponge collection

Information on sponge collection sites and dates are provided in Supplementary Table 1. For each specimen, one sample was subjected to metagenomic sequencing as described below (Extended Data Table 2).

Protocol A—enrichment of filamentous bacteria and DNA isolation

All protocol variants were applied to freshly collected sponges unless stated otherwise.

***T. swinhoei* WA and *T. swinhoei* WB.** The enrichment of filamentous bacteria and subsequent DNA isolation and sequencing were conducted in a previous study³⁰. The sequence dataset of that study was used for reanalysis as described below.

***Theonella* sp. 1 BA and *D. dissoluta*.** Filamentous bacteria were mechanically enriched before DNA isolation using a modified method reported by Bewley et al.²⁵. Sponge tissue (1 cm³) was soaked in 10 ml of calcium-free and magnesium-free artificial sea water (CMF-ASW; 10 mM Tris-HCl pH 8, 2.5 mM EGTA, 2.15 mM NaHCO₃, 33 mM Na₂SO₄, 9 mM KCl and 449 mM NaCl). After overnight incubation at 4 °C under gentle mixing, the tissue was cut into small pieces using a sterile scalpel and transferred to a new 15-ml conical tube. The sample was submerged in PBS (8.4 mM Na₂HPO₄, 1.5 mM KH₂PO₄ and 150 mM NaCl, pH 7.5, sterile-filtered and stored at room temperature), collagenase enzyme (1 µl per ml of PBS; final concentration: 240 µg ml⁻¹) was added and the mixture was incubated at 37 °C for 1 h. Subsequently, 10 ml of CMF-ASW was added and the sponge tissue was incubated at 4 °C for 2.5 h while mixing gently. After passing the sample through a 40-µm nylon filter into a 50-ml conical tube, the retained sponge tissue was transferred to a sterile mortar and ground with a pestle. The ground sample was then filtered through another 40-µm nylon filter into a new 50-ml conical tube and the filter was washed with 10–15 ml of CMF-ASW. The filtrates were combined and centrifuged for 10 min at 700g to sediment tissue and bacterial cells. The supernatant was carefully transferred into a new tube and the pellet was resuspended in 10 ml of CMF-ASW and centrifuged for 10 min at 20g to remove sponge tissues and unwanted debris. The supernatant was then transferred into a new 15-ml conical tube and the pellet was resuspended in 6 ml of CMF-ASW followed by another centrifugation step (10 min at 200g) to remove unicellular bacteria. The supernatant was again transferred to a new 15-ml conical tube and the pellet was resuspended again in 6 ml of CMF-ASW for another round of centrifugation (10 min at 200g) to further wash the now enriched filamentous bacterial cells. All centrifugation steps were performed at 4 °C. The cell fractions were assessed by microscopic analysis of each fraction. The DNA isolation was performed from enriched filamentous bacterial cells. For this, 1.2 ml of the 6 ml of enriched filamentous bacterial cells were used to pellet the filamentous bacteria (centrifugation for 3 min at maximum speed). The supernatant was removed and the pellet was resuspended in 250 µl of resuspension buffer (30 mM Tris-HCl pH 8.0, 1 mM EDTA and 0.1% SDS) supplemented with 15 µl of proteinase K solution (20 mg ml⁻¹). After incubation for 30 min at 50 °C, the treated cells were cooled on ice for 2 min followed by centrifugation at maximum speed for 5 min. The supernatant was extracted with one volume of phenol, chloroform

and isoamyl alcohol (25:24:1, v/v/v). The extraction was centrifuged at 9,000g for 10 min and the aqueous phase was transferred to a fresh tube. After another two rounds of extraction with one volume of chloroform, the aqueous phase was transferred to a precooled tube and DNA precipitation was performed by addition of one volume of cold 2-propanol and 0.1 volumes of 3 M sodium acetate and overnight incubation at –20 °C. The sample was centrifuged at maximum speed for 30 min and the supernatant was carefully removed. The resulting DNA pellet was washed twice with one volume of 70% ethanol (centrifugation at maximum speed for 10 min), followed by drying under a sterile hood for 5 min. The DNA was resuspended in 30 µl of elution buffer NE (Macherey-Nagel, NucleoSpin) and incubated overnight at 4 °C. To further remove RNA from the sample, the DNA solution (28 µL) was treated with 1.5 µl of RNase A (10 mg ml⁻¹) and incubated at 30 °C for 70 min. The treated DNA solution was then precipitated for 1 h at –20 °C with isopropanol and sodium acetate (1:0.1, v/v) following the same procedure as described above for washing the DNA and subsequent dilution of the DNA pellet overnight at 4 °C. The quality of the DNA was assessed by absorption at 230 nm, 260 nm and 280 nm and by gel electrophoresis.

Protocol B—metagenomic DNA extraction from sponge samples

***A. cribrophora*.** The sponge was stored in RNAlater at –20 °C upon collection. The DNA of this sponge was isolated as previously reported by Peters et al.³⁹. In brief, the defrosted sponge was rinsed with ASW before it was minced and homogenized using a Precellys 24 homogenizer (Bertin). The DNA was then isolated using the standard protocol of the DNeasy PowerSoil pro kit (Qiagen). Additionally, high-molecular-weight (HMW) DNA was isolated for Oxford Nanopore sequencing using the MagAttract HMW DNA kit (Qiagen), following the ‘disruption/lysis of tissue’ protocol according to the manufacturer’s instructions, followed by the ‘manual purification of HMW genomic DNA from fresh or frozen tissue’ protocol. Sponge pieces were weighed after thawing and squeezed to remove RNAlater but were not rinsed. Samples were treated following the set of standard protocols mentioned above, except gentle mixing was used instead of vortexing and only a Pipetman P1000 pipette was used to handle the DNA.

***T. swinhoei* YB, *D. calyx* and *D. kiiensis*.** Single-bacterial genome sequencing of ‘*Entotheonella*’ was conducted as previously described⁴⁶. In short, the sponge tissue was minced in CMF-ASW and the fraction that passed through a 40-µm mesh was collected as the bacterial fraction. Then, filamentous bacteria were enriched by centrifugation at 1,000g. After 30 s, the supernatant was collected and the pellet was isolated at 10 min. Sponge tissue or unicellular organisms were removed by each step. Filamentous bacteria were suspended in PBS and lysed by Ready-lyse lysozyme (Epicentre; 10 U per µl, 37 °C, 30 min), proteinase K (Promega; 1 mg ml⁻¹, 50 °C, 30 min) and heat treatment (95 °C, 15 min). The DNA was purified with a DNeasy blood and tissue kit (Qiagen) from the lysate.

Protocol C—acquisition of single-amplified genomes (SAGs)

***T. swinhoei* YB, *D. calyx*, *D. kiiensis* and *Theonella* sp. 2 BT.** Single-bacterial genome sequencing of ‘*Entotheonella*’ was conducted as previously described⁴⁶. In short, the sponge tissue was minced in CMF-ASW and the fraction that passed through a 40-µm mesh was collected as the bacterial fraction. Then, filamentous bacteria were enriched by centrifugation at 1,000g. After 30 s, the supernatant was collected and the pellet was isolated at 10 min. Sponge tissue or unicellular organisms were removed by each step. Filamentous bacteria were suspended in PBS and encapsulated into microdroplets with a diameter of 50 µm (ref. 46). The droplets containing single ‘*Entotheonella*’ filaments were manually picked with a micropipette (Drummond) under microscopic observation and isolated into 0.2-ml PCR tubes. The isolated

bacteria were lysed by Ready-lyse lysozyme (Epicentre; 10 U per μ l, 37 °C, 30 min), proteinase K (Promega; 1 mg ml⁻¹, 50 °C, 30 min) and heat treatment (95 °C, 15 min). To acquire single-bacterial amplified ‘Entotheonella’ genomes, multiple displacement amplification (MDA) was performed for 3 h with the REPLI-g single-cell kit (Qiagen). The MDA reactions were performed with 40 single filaments each from *D. calyx* and *D. kiiensis* and 96 single filaments each from *T. swinhoei* YB and *Theonella* sp. 2 BT.

DNA sequencing

***T. swinhoei* WA and *T. swinhoei* WB.** The isolated metagenomic DNA from the fraction of enriched filamentous bacteria was sequenced in a previous study³⁰ and was here subjected to an improved binning analysis as described below.

***Theonella* sp. 1 BA and *D. dissoluta*.** The isolated metagenomic DNA from the enriched filamentous bacterial fraction was sequenced by the Functional Genomics Center Zürich using an Illumina HiSeq2500 system.

***A. criobophora*.** The metagenomic DNA samples isolated from this sponge were sequenced by Novogen Europe using the Illumina Novaseq600 platform and the PE150 library and sequencing kits³⁹. Additionally, the extracted HMW DNA was sequenced in two rounds using an Oxford Nanopore Technologies MinION Mk1C. For the first round, the ligation sequencing kit SQK-LSK109 (Oxford Nanopore Technologies), the NEBNext Companion Module for the Oxford Nanopore Technologies ligation sequencing kit (New England Biolabs) and NBD104 barcodes (Oxford Nanopore Technologies) were used to make the sequencing libraries, following the ‘ligation sequencing gDNA–native barcoding’ (SQK-LSK109 with EXP-NBD104) protocol. The second sequencing library was made using the same kits, except new barcodes from the NBD114 kit (Oxford Nanopore Technologies) were used and the ‘ligation sequencing gDNA–native barcoding’ (SQK-LSK109 with EXP-NBD104 and EXP-NBD114) protocol was followed. Then, 200-ng samples of five sponge libraries were combined for the final sequencing library. The first round of sequencing was performed on an already used, flushed flowcell (R9.4.1), with approximately 590 pores available. The second round of sequencing was performed on a new flowcell (R9.4.1), with approximately 1,301 pores available.

***T. swinhoei* YB, *D. calyx*, *D. kiiensis* and *Theonella* sp. 2 BT.** Sequencing libraries were prepared from each SAG using the Nextera XT kit and short-read sequencing with MiSeq (Illumina) was conducted. Additionally, sequencing libraries were prepared with the rapid sequencing kit (Nanopore) from metagenomic DNA of the isolated filamentous bacterial fraction and sequenced by MinION (Nanopore) using the flowcell R9.4.1. Regardless of the genome construction method, at least nine MDA products were used to determine the genome of each variant.

Assembly and binning

***T. swinhoei* WA and *T. swinhoei* WB.** DNA sequencing, assembly and binning was performed in a previous study³⁴. For the work on *T. swinhoei* WB conducted in that study, the binning generated a single, inseparable bin containing the genomes of two ‘*Ca. Entotheonella*’ genomes at near-identical coverage. The assembly of the ‘*Ca. E.serta*’ TSWA1 single-bacterial genome from *T. swinhoei* WA³⁴ now allowed a refinement of this bin into sequences of very high identity (‘*Ca. E.serta*’ TSWB1) and moderate identity (‘*Ca. E. consors*’ TSWB2) and a small fraction with no apparent homology (unknown source). The latter was discarded.

***Theonella* sp. 1 BA and *D. dissoluta*.** The raw reads were assembled using SPAdes and binned on the basis of tetranucleotide frequency

and sequence coverage in a process described in more detail below. For the quality control of the metagenomes, BBduk (version 37.55, Joint Genome Institute) was first used in right-trimming mode with a *k*-mer length of 23 down to 11 and a Hamming distance of 1 to filter out sequencing adaptors. A second pass with a *k*-mer length of 31 and a Hamming distance of 1 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 14 and an average quality score cutoff of 20, with reads under 45 bp or containing Ns subsequently rejected. When a metagenomic assembly required more than 3 TB of RAM to complete, the reads were first *k*-mer-normalized with BBnorm (version 37.55, Joint Genome Institute) using a minimum depth of 2 and target depth of 80. The normalized paired-end and unnormalized singleton reads of each read set were assembled using metaSPAdes⁹⁹ (version 3.11.0) without the error correction module but otherwise default parameters. Scaffolds smaller than 1 kbp were then filtered out. For the binning, the quality-controlled paired-end reads were aligned to the assembled scaffolds using BWA (version 0.7.17)¹⁰⁰ and then filtered with a Python script for an identity of at least 97%, an alignment length of 200 bp and a minimum alignment coverage of 90% of the read length. The alignments were then sorted by SAMtools (version 1.9)¹⁰¹. Coverage depth across the scaffolds was calculated using the MetaBAT2 (version 2.12.1)¹⁰² jgi_summarize_bam_contig_depths script and this information was then used by MetaBAT2 to bin the scaffolds with default parameters.

***A. criobophora*.** Metagenomic reads were processed for quality trimming as described in Peters et al.³⁹. Using BBtools suite (version 37.64) with parameters *ktrim* = *r*, *k* = 23, *mink* = 7, *hdist* = 1, *tpe*, *tbo*, *qtrim* = *rl*, *trimq* = 20, *ftm* = 5, *maq* = 20 and *minlen* = 50, adaptors were removed and quality filtering and normalization were performed. The short and long reads were used for a hybrid assembly using metaSPAdes (version 3.12) with the ‘--only-assembler’ flag. Binning was performed using MetaWRAP (version 1.2) with minimum completeness of 50% and maximum contamination of 10%, as in Peters et al.³⁹.

***T. swinhoei* YB, *D. calyx* and *D. kiiensis*.** The quality of acquired short reads for each SAG was controlled with fastp (version 0.20.0)¹⁰³ (options: -q 25 -r -x) and de novo assembly with SPAdes (version 3.12.0)¹⁰⁴ (option: --sc-careful) was implemented. Then, the taxonomy and genome completeness of the SAG contigs were evaluated with CheckM (version 1.0.6)⁴⁷ and QUAST (version 4.5)¹⁰⁵. On the basis of the ccSAG¹⁰⁶ method, strain-level clustering of ‘Entotheonella’ SAGs was implemented. Then, ‘Entotheonella’ long reads were extracted by short-read mapping with SAGs of each ‘Entotheonella’ strains. Lastly, draft genomes of ‘Entotheonella’ were acquired by de novo assembly of long reads by Canu (version 1.4)¹⁰⁷ and polished by Pilon (version 1.22)¹⁰⁶ using short reads of the same strain. For *T. swinhoei* YB, the raw reads generated in the study by Wilson et. al.²⁹ were merged with the newly generated single-bacterial assembled data to produce a hybrid genome.

***Theonella* sp. 2 BT.** The quality of acquired short reads for each SAG was controlled with fastp (version 0.20.0)¹⁰³ (options: -q 25 -r -x) and de novo assembly with SPAdes (version 3.12.0)¹⁰⁴ (option: --sc-careful) was implemented. Then, the taxonomy and genome completeness of the SAG contigs were evaluated with CheckM (version 1.0.6)⁴⁷ and QUAST (version 4.5)¹⁰⁵. Finally, strain-level clustering of ‘Entotheonella’ SAGs and coassembly of clustered SAGs were implemented on the basis of the ccSAG¹⁰⁶ method.

Additional genome processing

The quality of the bins was assessed using the CheckM (version 1.0.13)⁴⁷ lineage workflow, which included taxonomic assignment and the generation of summary plots. Bins with $\geq 90\%$ completeness and $\leq 5\%$

contamination were deemed of high quality, those with $\geq 70\%$ completeness and $\leq 10\%$ contamination were deemed of good quality, those with $\geq 50\%$ completeness and $\leq 10\%$ contamination were deemed of medium quality and any bins with $< 50\%$ completeness or $> 10\%$ contamination were deemed of low quality. Genes were predicted with Prodigal (version 2.6.3)¹⁰⁸ in meta mode (-p meta) with the closed end (-c) and mask Ns (-m) options. Contigs were taxonomically identified with Kaiju (version 1.6.2)¹⁰⁹ against a provided subset of the National Center for Biotechnology Information BLASTnr database containing all proteins belonging to archaea, bacteria, viruses, fungi and microbial eukaryotes (nr_euk).

Contigs that were independently sequenced in other studies^{29,34,36,37} were manually scaffolded if applicable using published sequence data. Where SAG data were available that corresponded to phylotypes also identified in metagenomic samples (*T. swinhoei* WB and *Theonella* sp. 1BA), BLAST searches were used to retrieve additional nonbinned contigs where they could be unambiguously assigned to a draft MAG. After genome assembly, all genomes were scaffolded using the Multi-CSAR¹¹⁰ reference-based contig scaffolder, using each other as references, resulting in a notable reduction in the number of contigs and improved NS0 and LS0 values (Supplementary Table 2). Contigs below 500 bp in length were removed.

Evaluation of sequence quality was performed using the lineage_wf command in CheckM⁴⁷. Phylogenomic analysis was performed using the online tool autoMLST⁵⁰. The default nearest organisms and default MLST genes were selected, IQ-TREE Ultrafast Bootstrap was performed analysis with 1,000 replicates, ModelFinder was run, inconsistent MLST genes were filtered, the fast alignment mode (MAFFT FFT-NS-2) was implemented and a concatenated alignment was created. FastANI⁴⁹ analysis was performed using the comparison mode 'many to many' and a matrix was created (Supplementary Fig. 3). Annotation of genes was performed with RASTtk¹¹¹.

Tree-building methods

Automated multilocus species tree. The autoMLST tree in Fig. 2b was generated using the autoMLST tool (<https://automlst.ziemertlab.com/index>)⁵⁰ in de novo mode. The pipeline selects single-copy genes present in the organism set and infers a tree from the extracted sequences. The default nearest organisms and default MLST genes were selected, IQ-TREE Ultrafast Bootstrap analysis (1,000 replicates) was performed, ModelFinder was run, inconsistent MLST genes were filtered, fast alignment mode (MAFFT FFT-NS-2) was implemented and a concatenated alignment was generated. To run a larger dataset, the closest genomes to the symbiont genome of *A. cribophora* were identified with GTDB-Tk⁵¹ and used as reference genomes to run autoMLST locally and generate Supplementary Fig. 2. The same options were used including filtering of genes with inconsistent phylogeny (as described in the autoMLST methods) to safeguard against possible contamination. A list of the selected genes and functions can be found in Supplementary Table 12.

16S rRNA gene analysis and tree-building methods. 16S rRNA genes or fragments were identified using the ssu_finder method incorporated in CheckM⁴⁷. The sequences were then aligned using MUSCLE Alignment in Geneious 8.1.9 and the following settings: maximum number of iterations, 16; optimization, diagonal (keep tree from iteration 1; distance measure in iteration 1, kmer4_6; clustering method in iterations 1 and 2, UPGMB; tree-rooting method in iterations 1 and 2, pseudo; sequence weighting scheme in iterations 1 and 2, CLUSTALW; terminal gaps, half penalty; anchor spacing, 32; diagonals, min length = 24; minimum column anchor scores, min best = 90; hydrophobicity, multiplier = 1.2); optimization, anchor (keep tree from iteration 2; distance measure in subsequent, pctif_kimura; clustering method in subsequent, UPGMB; tree-rooting method in subsequent, pseudo; sequence weighting scheme in subsequent, CLUSTALW; objective score, spm;

gap open score = -1; diagonals, margin = 5; minimum column anchor scores, min smoothed = 90; hydrophobicity, window size = 5).

To infer trees, only complete genes (more than 1,500 nt) were taken into account for comparison to the complete gene deposited for '*Ca. E. palauensis*' (AF130847.1). The trees were generated in MEGA7 (ref. 112) using the neighbor-joining and maximum-likelihood methods. More details can be found in the figure caption of Supplementary Fig. 5.

Construction of the phylogenetic tree for the Cmb terpene synthase.

The tree was inferred through the tree builder function of Geneious 8.1.9 (alignment type, global alignment with free end gaps; cost matrix, Blosum45; genetic distance model, Jukes–Cantor; tree-building method, neighbor joining; gap open penalty, 8; gap extension penalty, 2).

Biochemical studies

BGC analysis. To evaluate the biosynthetic potential of the analyzed genomes, antiSMASH⁷⁵ was run on all genomes using the webserver (<https://antismash.secondarymetabolites.org/#!/start>) in relaxed mode and with all extra features selected (KnownClusterBlast, ClusterBlast, SubClusterBlast, MIBiG cluster comparison, ActiveSiteFinder, RREfinder, Cluster Pfam analysis, Pfam-based GO term annotation, TIGRFam analysis and TFBS analysis). All the gbks files from the antiSMASH output were combined and analyzed using BiG-SCAPE with the following parameters: -v --mode auto --mibig21 --mix --cutoffs 0.5 --include_singletons. The output was further processed with Cytoscape (version 3)¹¹³. Here, the network file for the group labeled mix was used for further visualization. The nodes were divided into biosynthetic groups of thiotemplate-based pathways, RiPPs, terpenes and other. The latter contained all BGCs for indoles, ladderanes, ectoine, phosphonates and homoserine lactones but not for thiotemplate-based, RiPP or terpene biosynthesis. Furthermore, BGCs were manually analyzed for reoccurring modular architectures within the different genomes and occurrence of protein families on the basis of automated gene annotations generated with RASTtk and BLAST searches.

Isolation of discodermins. Frozen sponge specimens of *D. kiiensis* (100 g, wet weight) were extracted with methanol and the extract was partitioned between *n*-butanol and water. The *n*-butanol fraction was then fractionated by octadecylsilyl flash chromatography (C18-prep, Nacalai Tesque, Japan) using a stepwise gradient with H₂O and methanol (0% to 100% methanol). The 80% methanol fraction was further purified by HPLC (Cosmosil MS-II, 10 × 250 mm; Nacalai Tesque) using CH₃CN:H₂O (2:3) containing 0.05% trifluoroacetic acid. Finally, another round of HPLC, using the same column with CH₃OH:H₂O (7:3) containing 0.05% trifluoroacetic acid, was performed to obtain discodermin A (10.2 mg), discodermin B (4.5 mg) and discodermin D (3.4 mg).

Overexpression of dscE. The gene *dscE* was codon-optimized for *E. coli* and synthesized by Twist Bioscience with an N-terminal His₆ tag and within the pET-28a(+) backbone. Electrocompetent NiCo21(DE3) *E. coli* cells (New England Biolabs) were transformed with the pET-28a(+)-DscE plasmid (KanR) together with the plasmids pDB1282 (AmpR)⁷⁷ and pBAD42-BtuCEDFB (SpecR)⁷⁸. Precultures were prepared in Luria–Bertani (LB) medium supplemented with the appropriate antibiotics and incubated overnight at 37 °C and 180 rpm. Then, 500 ml of terrific broth (TB) medium supplemented with 50 ml of glycerol, appropriate antibiotics, 0.5 mM δ -aminolevulinic acid, 300 mM CoCl₂ and 1.5 mM MeCbl were inoculated with 1% preculture and incubated at 37 °C and 180 rpm. At an optical density at 600 nm (OD₆₀₀) of 0.3, a 20-ml aliquot was taken (non-induced); protein expression of pDB1282 and pBAD42-BtuCEDFB was induced with 0.2% (w/v) L-arabinose and the medium was supplemented with 50 mM ammonium iron(II) sulfate and 300 mM L-cysteine. The culture was further incubated at 37 °C and 180 rpm until an OD₆₀₀ \approx 1.0 was reached (half-induced). After cooling the culture at 4 °C for 30 min,

protein expression of pET-28a(+)-DscE was induced with 1 mM IPTG. The culture was incubated overnight at 16 °C and 140 rpm.

Purification of DscE. After an aliquot was taken (induced), cells from the overexpression culture were harvested by centrifugation at 6,000g for 20 min and 4 °C. The pellet was transferred into an anaerobic chamber and dissolved in 1 ml of lysis buffer (50 mM HEPES pH 7.8, 300 mM KCl, 0.05% (v/v) Triton X-100, 10% glycerol and 20 mM imidazole) per 0.1-g pellet. The cells were sonicated four times for 2 min each at 20% amplitude, alternating between 5 s on and 5 s off (total lysate). The total lysate was cleared by centrifugation for 5 min at 14,000g and aliquots of the supernatant and the pellet were taken. The supernatant was incubated with 1 ml of Protino Ni-NTA agarose (Macherey-Nagel) for 1 h at 4 °C. After transferring the suspension onto an appropriate column, the flowthrough was collected. The sample was washed with 10 ml of lysis buffer and DscE was eluted by adding 5 ml of elution buffer (50 mM HEPES pH 7.8, 300 mM KCl, 0.05% (v/v) Triton, 10% glycerol and 100 mM imidazole) onto the column (eluate). The eluate was incubated with 1 ml of chitin resin (New England Biolabs) for 15 min at room temperature. After transferring the suspension onto an appropriate column, the flowthrough was collected (chitin eluate) and the buffer was exchanged with reconstitution buffer (50 mM HEPES pH 7.5, 300 mM KCl, 10% glycerol and 1 mM DTT) using a 30-kDa Amicon Ultra-0.5 centrifugal filter device (Merck) followed by a concentration step (concentrated and exchanged). Using all collected aliquots, a 12% SDS-PAGE was performed (Supplementary Fig. 19). The ultraviolet (UV)-visible light spectrum ($\lambda = 200\text{--}100\text{ nm}$) was recorded for the concentrated sample (Supplementary Fig. 20) and the protein concentration was measured using the UV absorbance at 280 nm and the calculated extinction coefficient ϵ .

Fe-S cluster reconstitution of DscE. The iron-sulfur clusters and cobalamin cofactors were reconstituted overnight at 4 °C. To approximately 127 mM DscE (one equivalent), 12 equivalents of L-cysteine, 13 equivalents of ammonium iron(II) sulfate, 20 equivalents of DTT, 2 equivalents of MeCbl, 1 mM IscS and 315 mM pyridoxalphosphate were added.

DscE in vitro assay. The reconstitution reaction was added to a 1-ml bed volume of TALON metal affinity resin (Takara) and incubated for 15 min at room temperature. After transferring the suspension onto an appropriate column, the flowthrough aliquot was collected. The sample was washed with 10 ml of lysis buffer and collected in three fractions (aliquots W1, W2 and W3). Proteins were eluted with 5 ml of elution buffer and collected in three fractions as well (aliquots E1, E2 and E3). Using these seven aliquots, SDS-PAGE analysis was performed using a 12% SDS-PAGE gel (Supplementary Fig. 19). As most of the DscE enzymes were in the flowthrough, this fraction was concentrated using an equilibrated 30-kDa Amicon Ultra-0.5 centrifugal filter device. Furthermore, the buffer was exchanged to reaction buffer (50 mM HEPES pH 7.5, 150 mM KCl and 10% glycerol). After this buffer exchange, the UV-visible light spectrum ($\lambda = 200\text{--}100\text{ nm}$) was recorded for the concentrated sample (Supplementary Fig. 20) and the protein concentration was measured using the UV absorbance at 280 nm. In vitro reactions were set up as follows in a total volume of 50 μl in reaction buffer and incubated overnight at room temperature: 20 mM DscE, 0.2 mM SAM, 0.2 mM methyl viologen, 0.2 mM NADPH, 0.2 mM substrate, 0.1 mM MeCbl and 2 mM DTT. The following controls were included: (1) heat-inactivated DscE instead of DscE; (2) DMSO instead of substrate; and (3) no addition of reduction system (methyl viologen, NADPH and DTT) (Supplementary Fig. 23). The next day, reactions were transferred out of the anaerobic chamber and quenched by adding 50 μl of 0.5 M formic acid in methanol. Samples were analyzed using HPLC-MS with the following parameters: solvent A, $\text{H}_2\text{O} + 0.1\%$ formic acid; solvent B, $\text{CH}_3\text{CN} + 0.1\%$ formic acid; column, Kinetex

2.6 mm XB-C18 100 Å ($150 \times 4.6\text{ mm}$); flow rate, 1 ml min^{-1} ; column oven, 50 °C. The gradient was adjusted as follows: starting condition of 10% solvent B for 2 min, followed by a linear gradient over 10 min to 65% solvent B and an even steeper gradient over 1 min toward 98% solvent B. The column was further flushed with 98% solvent B for 3.5 min followed by a 0.4-min equilibration step to 10% solvent B. Before a new measurement, an equilibration step of 10% B over 3 min was performed. The MS instrument was operated in positive ionization mode at a scan range of 200–2,000 m/z and a resolution of 70,000. The spray voltage was set to 3.5 kV, S-lens was set to 50, sheath gas was set to 57.50, probe heater temperature was set to 462.50 °C and capillary temperature was set to 281.25 °C. The reaction was performed in triplicate and was repeated multiple times with freshly purified enzyme. Data analysis was conducted with Xcalibur 4.1 (Thermo Fisher).

Overproduction and purification of Cmb variants. The genes encoding CmbA homologs in ‘*Ca. E. sarta*’ TSWB1 and ‘*Ca. E. mit-sugo*’ TYSB3 were PCR-amplified from the synthetic genes (synthesized by Twist Bioscience; Supplementary Table 10) using the primers Cmb^{Es}-F and Cmb^{Es}-R or Cmb^{Em}-F and Cmb^{Em}-R, respectively (Supplementary Table 13). The PCR-amplified genes were analyzed on an agarose gel and the gene fragments were purified from there. Subsequent digestion with NdeI and HindIII and ligation into a pET28b(+) (Novagen) followed by introduction into *E. coli* DH5 α resulted in the plasmids pET28b(+)-Cmb^{Es} and pET28b(+)-Cmb^{Em}. The final plasmid constructs were then used to transform *E. coli* BL21 (DE3) (Stratagene).

E. coli BL21 (DE3) cells harboring pET28b(+)-Cmb^{Es} or pET28b(+)-Cmb^{Em} were precultured in LB medium containing 50 $\mu\text{g ml}^{-1}$ kanamycin at 37 °C. The preculture was used to inoculate TB medium containing 50 $\mu\text{g ml}^{-1}$ kanamycin and the cultures were grown at 37 °C for 2 h. Gene expression was then induced by the addition of IPTG at a final concentration of 0.1 mM and growth was continued at 18 °C for 14–16 h. The cells were harvested by centrifuging at 3,910g and 4 °C for 10 min. The cell pellets were resuspended in 50 mM Tris-HCl pH 8.0, 0.5 mM NaCl, 20 mM imidazole and 20% glycerol and cells were disrupted using a Branson Sonifier 250 (Emerson). The lysate was centrifuged at 34,700g at 4 °C for 10 min. The recombinant Cmb proteins were purified from the resulting supernatant using Ni-NTA Superflow resin (Qiagen). After washing with the buffer containing 50 mM Tris-HCl pH 8.0, 0.5 mM NaCl, 20 mM imidazole and 20% glycerol, the protein was eluted with 50 mM Tris-HCl pH 8.0, 0.5 mM NaCl, 250 mM imidazole and 20% glycerol. Finally, the protein was concentrated to an appropriate concentration with Vivaspin 10,000-kDa molecular weight cutoff.

Cmb in vitro assays. The standard assay was performed at 30 °C in a 100- μl reaction mixture containing 50 mM HEPES-NaOH pH 7.5, 2.0 mM MgCl_2 , 5.0 μM Cmb and 1.0 mM GGPP, GPP or FPP. The reaction was quenched by addition of 200 μl of ethyl acetate. After centrifugation, the upper layer was analyzed by GC-MS using a Shimadzu GCMS-QP2020. Sample introduction was performed by split injection onto a Shimadzu GLC SH-Rxi-5ms (5% diphenyl-95% dimethylpolysiloxane) column (30 m, 0.25-mm inner diameter, 0.25- μm film thickness). The injector temperature was 230 °C. The initial column temperature was 50 °C and this temperature was held for 1 min after injection. Next, the temperature was increased to 150 °C at $10\text{ }^\circ\text{C min}^{-1}$ and then to 280 °C at $20\text{ }^\circ\text{C min}^{-1}$. The temperature was held at 280 °C for the remainder of the 22.5-min program. Data analysis was conducted with LabSolutions CS (version 4.42).

Isolation and structure elucidation of cembrene A (21). To obtain 21, the Cmb^{Es} assay was performed in a 50-ml reaction mixture containing 50 mM HEPES-NaOH pH 7.5, 0.5 mM GGPP, 2.0 mM MgCl_2 and 2.9 μM Cmb^{Es}. The reactions were incubated at 30 °C overnight and extracted with hexane ($2 \times 100\text{ ml}$) and ethyl acetate ($2 \times 100\text{ ml}$). The combined organic layer was dried with MgSO_4 , concentrated under

reduced pressure and the target diterpene was purified by column chromatography on silica gel with *n*-hexane to yield cembrene A (1, 0.8 mg). NMR spectra were recorded on a JEOL ECA-600 spectrometer operating at 600 MHz for ^1H and 150 MHz for ^{13}C nuclei. NMR data were analyzed using Delta 5.3 (JEOL). The optical rotations were recorded with a P-2100 polarimeter (JASCO) and compared to the reported values for (*R*)-cembrene (−12) (ref. 87) and (*S*)-cembrene (+19.5) (ref. 88).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data supporting the findings of this study are available within the main text and the Supplementary Information. DNA sequences were deposited to the European Nucleotide Archive under BioProjects [PRJEB80215](#) (all except for '*Ca. P. opulenta*' AC1) and [PRJEB59408](#) ('*Ca. P. opulenta*' AC1) with the following accession numbers: '*Ca. E. symbiotica*' BT01, [GCA_964656635](#); '*Ca. E. inquilina*' BT02, [GCA_964656685](#); '*Ca. E. melakyensis*' BT03, [GCA_964656765](#); '*Ca. E. catenata*' BT04, [GCA_964656715](#); '*Ca. E. armillaria*' DC1, [GCA_964656755](#); '*Ca. E. tacita*' DD1, [GCA_964656785](#); '*Ca. E. baccata*' DD2, [GCA_964656735](#); '*Ca. E. tertia*' DD3, [GCA_964656775](#); '*Ca. E. monilis*' DK1, [GCA_964656725](#); '*Ca. E. melakyensis*' TCBA1, [GCA_964656645](#); '*Ca. E. sarta*' TCBA2, [GCA_964656625](#); '*Ca. E. sarta*' TSWA1, [GCA_964656745](#); '*Ca. E. sarta*' TSWB1, [GCA_964656705](#); '*Ca. E. consors*' TSWB2, [GCA_964656675](#); '*Ca. E. factor*' TSYB1, [GCA_964656695](#); '*Ca. E. gemina*' TSYB2, [GCA_964656665](#); '*Ca. E. mitsugo*' TSYB3, [GCA_964656655](#); '*Ca. P. opulenta*' AC1, [GCA_965178525](#). The *dsc* BGC was deposited to MIBiG with accession number [BGC0003182](#). Other data related to this work (for example, HPLC–HRMS) are available from the lead contact upon request.

References

- Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* **27**, 824–834 (2017).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
- Kang, D. D., et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
- Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
- Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
- Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
- Walker, B. J., et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
- Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
- Hyatt, D., et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
- Menzel, P., Ng, K. L. & Krogh, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* **7**, 11257 (2016).
- Chen, K. T., Shen, H. T. & Lu, C. L. Multi-CSAR: a multiple reference-based contig scaffold using algebraic rearrangements. *BMC Syst. Biol.* **12**, 139 (2018).

- Brettin, T. 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).
- Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
- Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

Acknowledgements

We thank A. L. Vagstad for providing HPLC-purified SAM and IscC, the MOZ-04 IFREMER 2016 expedition and L. Corbari (Muséum National d'Histoire Naturelle) on board for the collection of *Theonella* sp. and P. McCarthy for his support with experiments on the *D. dissoluta* samples. Some of the sequencing was performed at the Functional Genomics Center Zurich (FGCZ). T.S. and T.K. are grateful for funding by the JSPS KAKENHI (22H05120 to T.K.), T.W. is grateful for funding by the JSPS KAKENHI (grant numbers JP21H02635 and JP22H05128), K.T. is grateful for funding by JSPS KAKENHI (grant number 16H06279, PAGES), J.P. and D.S. are grateful for funding from the European Union's Horizon 2020 research and innovation program under grant agreement 101000392 (MARBLES) and J.P. is grateful for funding by the Swiss National Science Foundation (grant numbers 205320_185077 and 205320_219638), the Gordon and Betty Moore Foundation (grant number 9204; <https://doi.org/10.37807/GBMF9204>) and ETH Zurich (research grant number ETH-21 18-2).

Author contributions

Conceptualization, M.D., A.B.S., T.S., J.K.B.C., H.T. and J.P. Methodology, M.D., M.K., A.B.S., T.S., A.L., C.M.M., M.A.S., H.Y., Y.Y., J.K.B.C., Y.E., K.T.C. and E.E.P. Software, C.F. and M.A. Validation, M.D., A.B.S., T.S., A.L. and M.A.S. Formal analysis, M.D., A.L., J.K.B.C., Y.N., M.A., C.R., J.K. and J.P. Investigation, M.D., M.K., A.B.S., T.S., A.L., C.M.M., M.A.S., J.K.B.C., H.Y., Y.Y., Y.E., K.T.C., E.E.P., C.R., K.T. and J.P. Resources, C.F., P.C., S.P., D.S., A.W., K.T., I.A., T.W., H.T. and J.P. Data curation, M.D., A.L., M.A.S., C.F., J.K.B.C. and J.P. Writing—original draft, M.D., A.B.S., T.S., A.L., H.T. and J.P. Writing—review and editing, M.D., M.K., A.B.S., T.S., A.L., C.M.M., M.A.S., J.K.B.C., T.K., P.C., D.S., H.T., T.W. and J.P. Resources, C.F., P.C., S.P., D.S., A.W., K.T., I.A., T.W., H.T. and J.P. Visualization, M.D., A.B.S., T.S., A.L. and J.P. Supervision, M.D., J.K.B.C., I.A., T.W., H.T. and J.P. Project administration, M.D., H.T. and J.P. Funding acquisition, J.K.B.C., T.K., I.A., T.W., H.T. and J.P.

Funding

Open access funding provided by Swiss Federal Institute of Technology Zurich.

Competing interests

The authors declare no competing interests.

Additional information

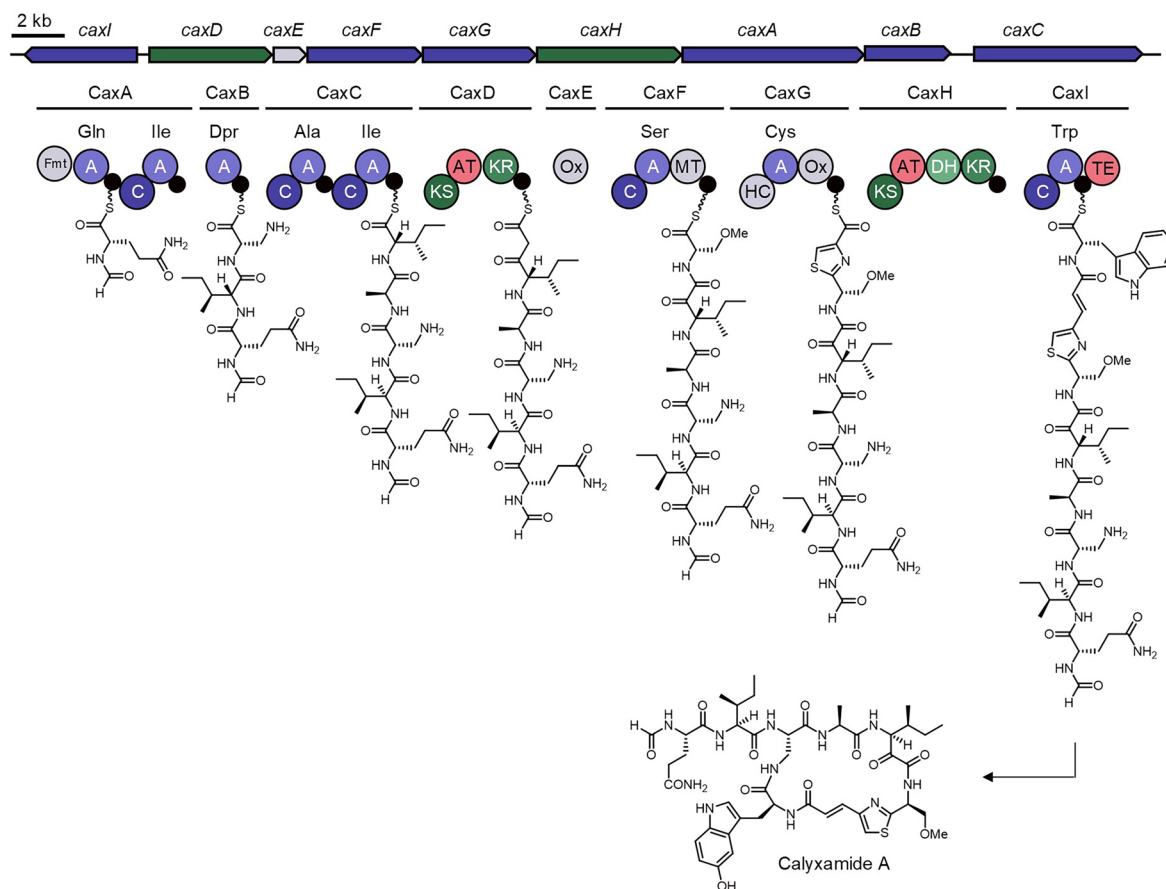
Extended data is available for this paper at <https://doi.org/10.1038/s41589-025-02066-0>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41589-025-02066-0>.

Correspondence and requests for materials should be addressed to Toshiyuki Wakimoto, Haruko Takeyama or Jörn Piel.

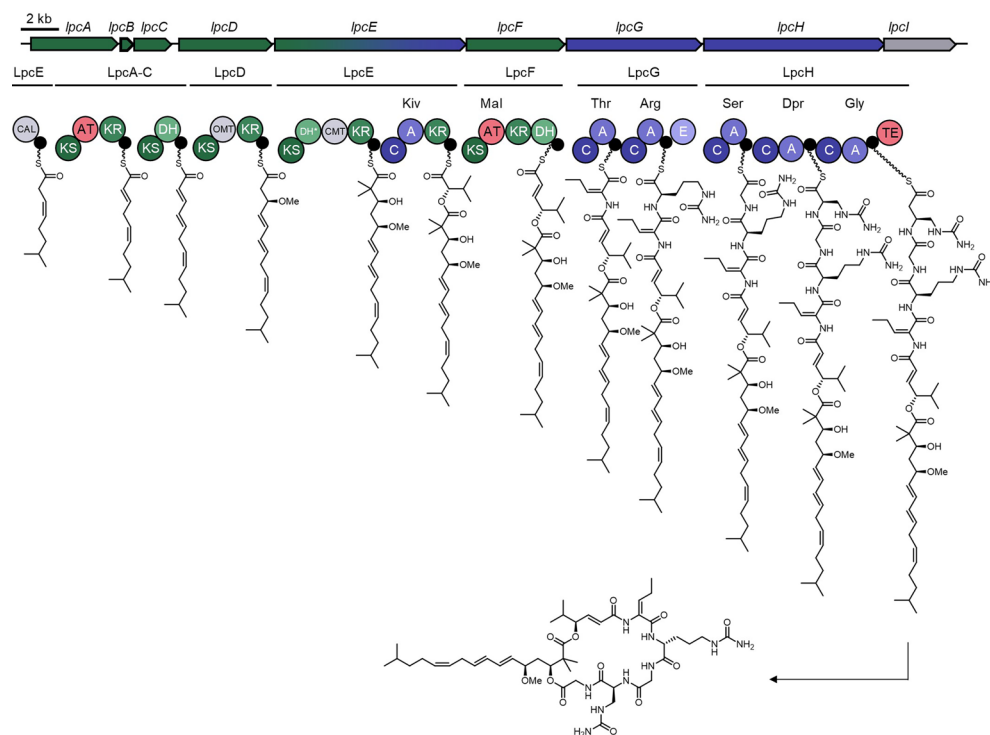
Peer review information *Nature Chemical Biology* thanks Zhiyong Li, Sarah Messenger, Gerardo Della Sala and the other, anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | Proposed biosynthetic model for calyxamide biosynthesis. The modular architecture of the enzymes is shown with intermediates attached to the carrier proteins (visualized as black circles). Labels above the adenylation domains refer to the substrate specificity

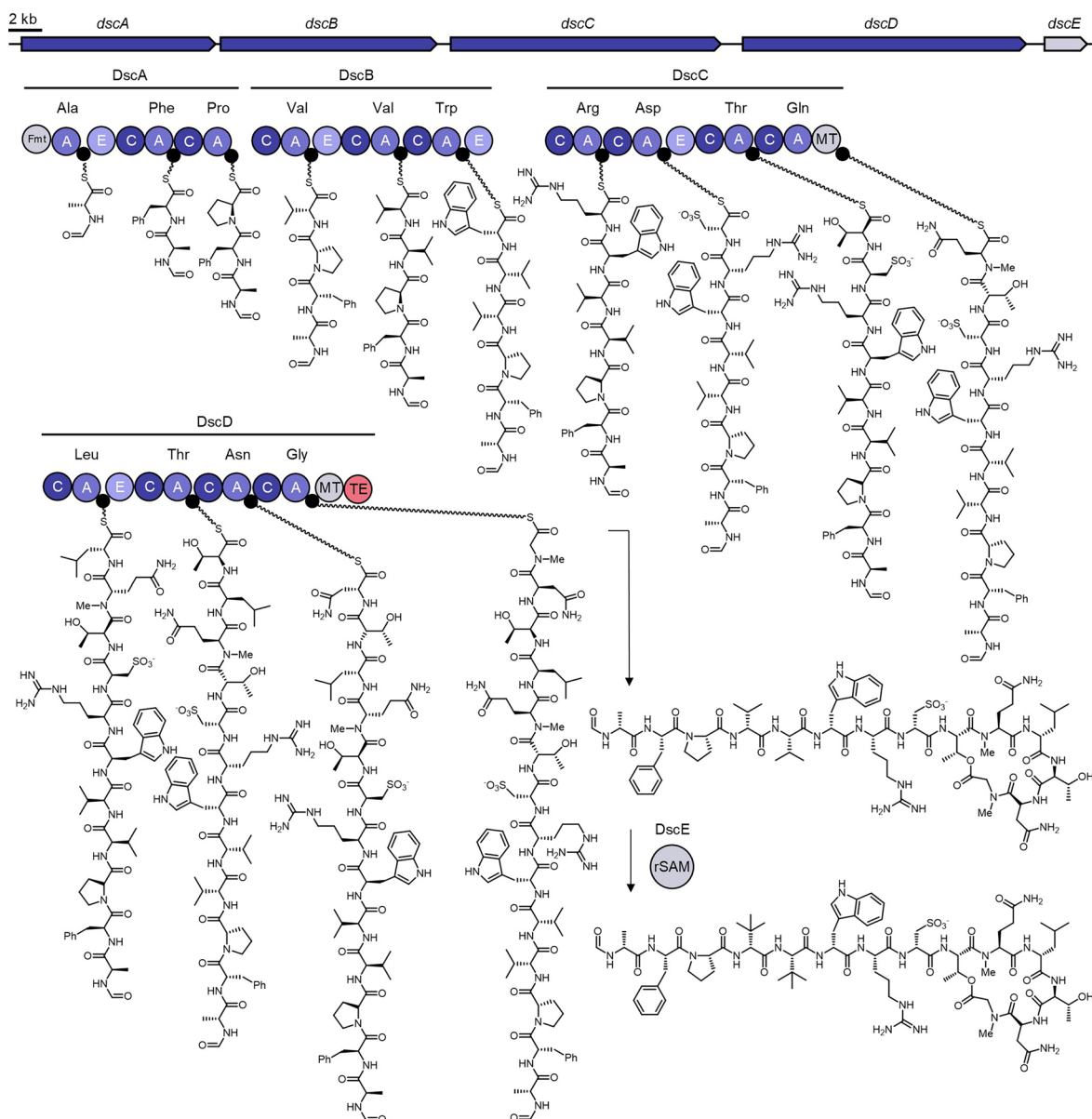
predicted by antiSMASH. Fmt, formyltransferase; A, adenylation domain; C, condensation domain; KS, ketosynthase; AT, acyltransferase; KR, ketoreductase; Ox, oxygenase; MT, methyltransferase; HC, heterocyclization domain; DH, dehydratase domain; TE, thioesterase; Dpr, 2,3-diaminopropionic acid.



Extended Data Fig. 2 | Proposed biosynthetic model for lipodiscamide biosynthesis. The modular architecture of the enzymes for lipodiscamide biosynthesis is shown with intermediates attached to the carrier proteins (visualized as black circles). Labels above the adenylation domains refer to the substrate specificity predicted by antiSMASH. CAL, coenzyme A ligase domain;

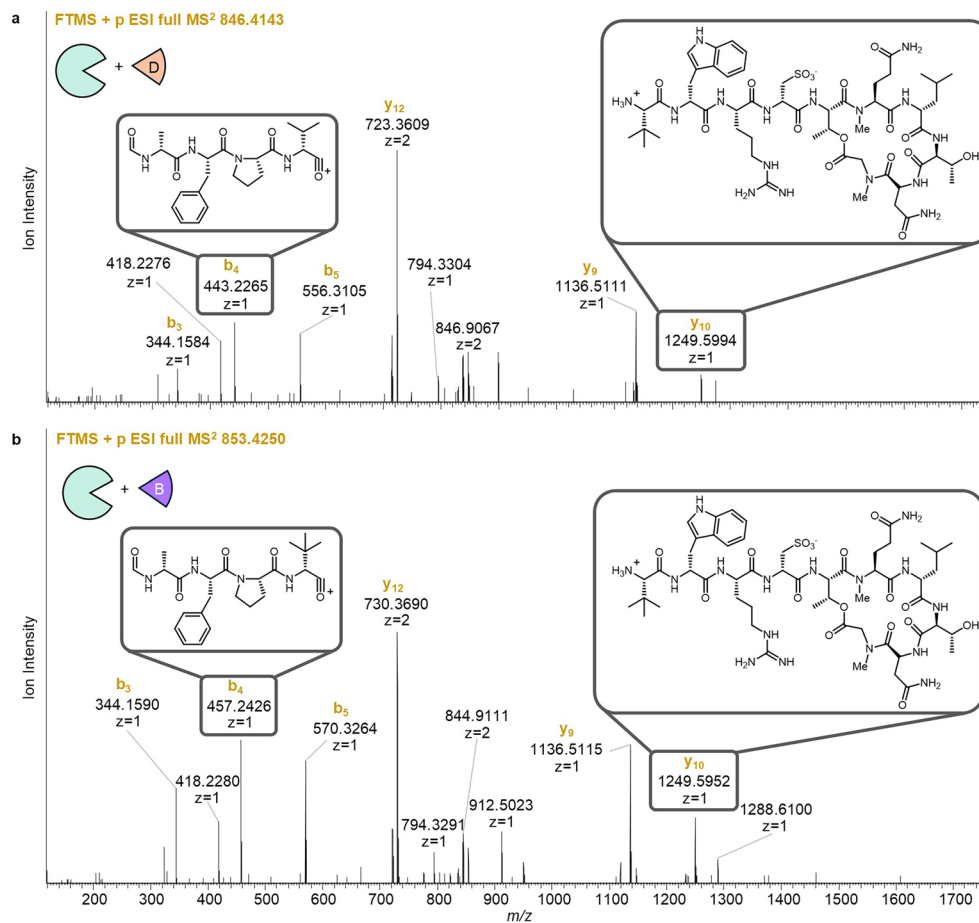
Lipodiscamide A

KS, ketosynthase; AT, acyltransferase; KR, ketoreductase; DH, dehydratase domain; OMT, *O*-methyltransferase; CMT, *C*-methyltransferase; C, condensation domain; A, adenylation domain; E, epimerization domain; TE, thioesterase; Kiv, α -keto isovaleric acid; Dpr, 2,3-diaminopropionic acid; * = putatively non-functional or degraded domain due to a shortened amino acid sequence.



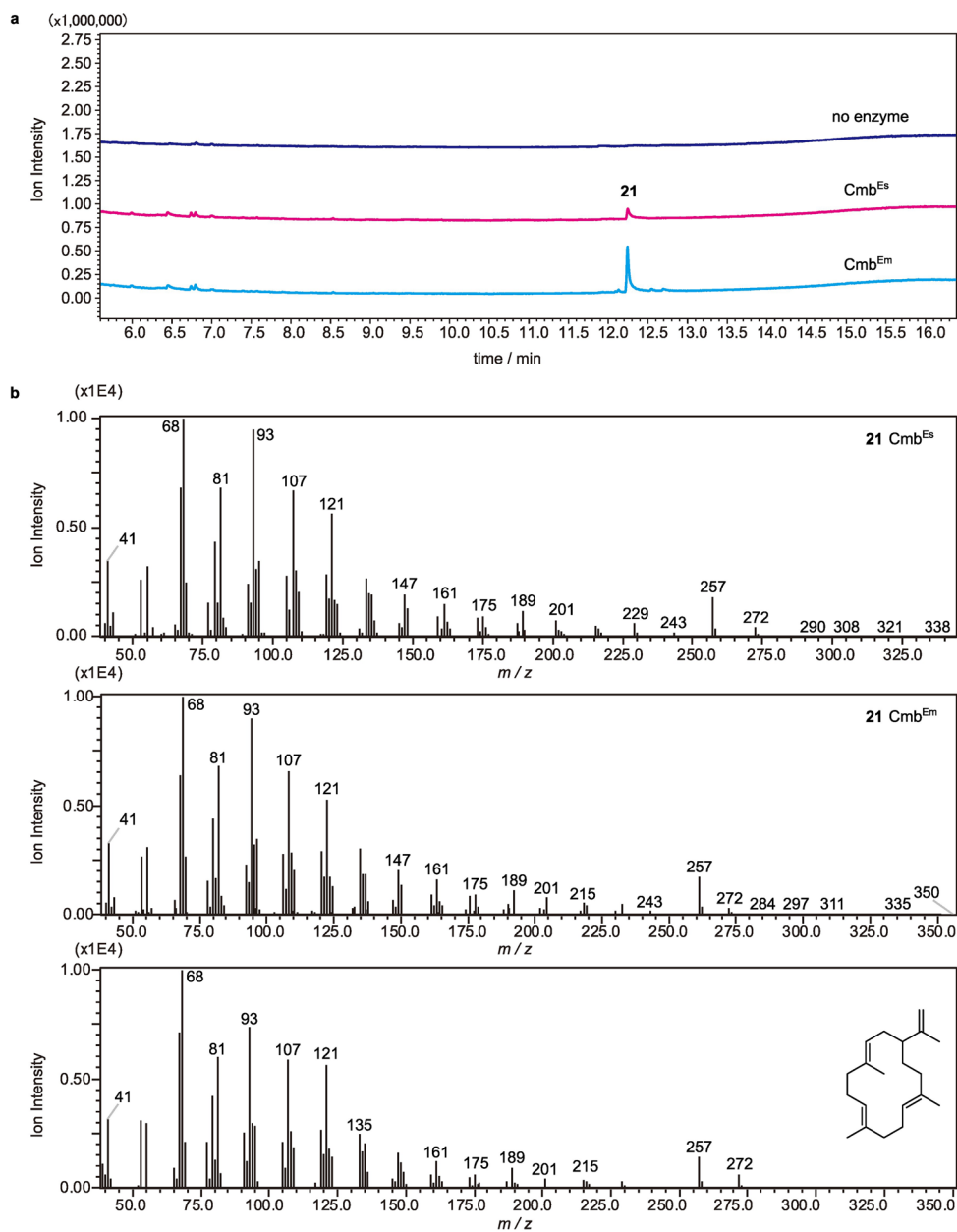
Extended Data Fig. 3 | Proposed biosynthetic model for discodermin biosynthesis. The modular architecture of the enzymes for discodermin biosynthesis is shown with intermediates attached to the carrier proteins (visualized as black circles). Labels above the adenylation domains refer to the

predicted substrate specificity by antiSMASH¹¹. Fmt, formyltransferase; A, adenylation domain; E, epimerization domain; C, condensation domain; MT, methyltransferase; TE, thioesterase; rSAM, SAM-dependent methyltransferase.



Extended Data Fig. 4 | MS² spectra of products from the *in vitro* reconstitution of rSAM DscE methyltransferase activity. The significant mass fragments for the respective product are depicted with potential fragments. **a**, Incubation of DscE with discodermin D (as visualized with the cartoon on the upper left).

The structures in the boxes are potential fragments that fit to the respective fragmentation m/z . **b**, Incubation of DscE with discodermin B (as depicted with the cartoon on the upper left). The structures in the boxes are potential fragments that fit to the respective fragmentation m/z .



Extended Data Fig. 5 | In vitro reconstitution of the terpene synthase Cmb with GGPP as a substrate. a, GC-MS TIC traces of ethyl acetate extracts of in vitro assays with FPP and terpene synthases. **b,** EI mass spectra of the reaction product of Cmb^{Es} (top), Cmb^{Em} (middle) and reference spectrum for cembrene A (bottom).

Extended Data Table 1 | NMR data for structure elucidation of compound **21**

Position	δC / ppm	δH / ppm		J	Hz
1	32.491	1.976	1.932	m, m	
2	124.107	5.176		t	7.4
3	134.93	-			
4	39.03	2.158	2.100	m, m	
5	24.97	2.254	2.158	m, m	
6	126.01	4.966		t	6.5
7	133.54	-			
8	39.50	2.033		m	
9	23.82	2.107		m	
10	121.87	5.047		t	6.5
11	134.00	-			
12	34.02	1.929	1.758	m, m	
13	28.20	1.654	1.356	m, m	
14	46.04	2.017		m	
15	149.39	-			
16	110.24	4.699	4.638	s	
17	19.39	4.646		s	
18	15.62	1.554		s	
19	15.40	1.575		s	7.4
20	18.13	1.546		overlapped	
Numbered structure					
					

^{13}C and ^1H NMR spectral data for structure elucidation of compound **21**. (^1H -NMR, 600 MHz; ^{13}C -NMR, 150 MHz; chloroform- d_6).

Extended Data Table 2 | Summary of methods on sponge sample processing

Sponge	Sponge work-up and applied DNA isolation protocol
<i>Theonella swinhoei</i> YB	Metagenomic DNA extraction from sponge samples and single-bacterial genomics method (Protocol B & C)
<i>Theonella swinhoei</i> WA	Enrichment of filamentous bacteria and single-bacterial genomics method (Protocol A & C)
<i>Theonella swinhoei</i> WB	Enrichment of filamentous bacteria and single-bacterial genomics method (Protocol A & C)
<i>Theonella</i> sp. 1 BA	Enrichment of filamentous bacteria (Protocol A)
<i>Theonella</i> sp. 2 BT	Single-bacterial genomics method (Protocol C)
<i>Discodermia calyx</i>	Metagenomic DNA extraction from sponge samples and single-bacterial genomics method (Protocol B & C)
<i>Discodermia kiensis</i>	Metagenomic DNA extraction from sponge samples and single-bacterial genomics method (Protocol B & C)
<i>Discodermia dissoluta</i>	Enrichment of filamentous bacteria (Protocol A)
<i>Aciculites cribrifera</i>	Metagenomic DNA extraction from sponge samples (Protocol B)

DNA isolation and prior work-up of sponge samples.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
<input type="checkbox"/>	<input checked="" type="checkbox"/> The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
<input type="checkbox"/>	<input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
<input checked="" type="checkbox"/>	<input type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
<input checked="" type="checkbox"/>	<input type="checkbox"/> A description of all covariates tested
<input type="checkbox"/>	<input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
<input checked="" type="checkbox"/>	<input type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
<input checked="" type="checkbox"/>	<input type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give <i>P</i> values as exact values whenever suitable.</i>
<input checked="" type="checkbox"/>	<input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
<input checked="" type="checkbox"/>	<input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
<input checked="" type="checkbox"/>	<input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	Sequencing techniques: Illumina HiSeq2500 system and Illumina Novaseq600 platform (Illumina), MinION Mk1C (Oxford Nanopore Technologies), MiSeq (Illumina); MS acquisition: Xcalibur 4.1, Chromeleon Xpress 7.2 (Thermo Fisher), HPLC: Open Lab CDS 2.2 (Agilent), NMR acquisition: Delta v5.3 (JEOL, Japan), GC-MS acquisition: LabSolutions CS (v.4.42) (Shimadzu)
Data analysis	DNA assemblies and binning: BBduk (v37.55, Joint Genome Institute), BBNorm (v37.55, Joint Genome Institute) (BBMap download SourceForge.net), metaSPAdes (v3.11.0 and v3.12) (https://doi.org/10.1101/gr.213959.116), SAMtools (v1.9), (https://doi.org/10.1093/gigascience/giab008), BWA (v0.7.1 and v0.7.17), (Burrows-Wheeler Aligner (sourceforge.net)), MetaBAT2 (v2.12.1 and 2.15), (https://doi.org/10.7717/peerj.7359), metaSPAdes (v3.12) (doi:10.1101/gr.213959.116), fastp (v0.20.0) (https://doi.org/10.1093/bioinformatics/bty560), Flye (v2.8.1-b1676), (https://doi.org/doi:10.1038/s41592-020-00971-x), Canu (v.1.4) (https://doi.org/10.1101/gr.215087.116), Pilon (v.1.22) (https://doi.org/10.1371/journal.pone.0112963), MetaWRAP v1.2, (https://doi.org/10.1186/s40168-018-0541-1), SPAdes 3.12.0, (https://doi.org/10.1002/cpb.102), CheckM v1.0.6 and v1.0.13 (10.1101/gr.186072.114), QUAST (v4.5), (QUAST (sourceforge.net), Kaiju (v.1.6.2) (https://doi.org/10.1038/ncomms11257), Prodigal (v2.6.3) (https://doi.org/10.1186/1471-2105-11-119), Multi-CSAR (https://doi.org/10.1186/s12918-018-0654-y), autoMLST (https://doi.org/10.1093/nar/gkz282), FastANI (https://doi.org/10.1038/s41467-018-07641-9), RASTtk (https://doi.org/10.1038/srep08365), GTDB-Tk (https://doi.org/10.1093/nar/gkab776), Geneious v.8.1.9 (Biomatters Limited), MEGA7 (https://doi.org/10.1093/molbev/msw054), antiSMASH7 (https://doi.org/10.1093/nar/gkad344), BiG-SCAPE (https://doi.org/10.1038/s41589-019-0400-9), Cytoscape 3 (https://doi.org/10.1101/gr.1239303), HPLC-HRMS: Xcalibur 4.1 (Thermo Fisher); NMR Topspin 4.1 (Bruker), GC-MS:LabSolutions CS (v4.42) (Shimadzu), NMR: Delta v5.3 (JEOL, Japan)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All data supporting the findings of this study are available within the main text and the supplemental information. DNA sequences are deposited at the European Nucleotide Archive (ENA) under the BioProject numbers PRJEB80215 (all except for 'Ca. P. opulenta' AC1) and PRJEB59408 ('Ca. P. opulenta' AC1), and accession numbers for assemblies are as following:

'Ca. E. symbiotica' BT01: GCA_964656635, 'Ca. E. inquilina' BT02: GCA_964656685, 'Ca. E. melakyensis' BT03: GCA_964656765, 'Ca. E. catenata' BT04: GCA_964656715, 'Ca. E. armillaria' DC1: GCA_964656755, 'Ca. E. tacita' DD1: GCA_964656785, 'Ca. E. baccata' DD2: GCA_964656735, 'Ca. E. tertia' DD3: GCA_964656775, 'Ca. E. monilis' DK1: GCA_964656725, 'Ca. E. melakyensis' TCBA1: GCA_964656645, 'Ca. E. sarta' TCBA2: GCA_964656625, 'Ca. E. sarta' TSWA1: GCA_964656745, 'Ca. E. sarta' TSWB1: GCA_964656705, 'Ca. E. consors' TSWB2: GCA_964656675, 'Ca. E. factor' TSYB1: GCA_964656695, 'Ca. E. gemina' TSYB2: GCA_964656665, 'Ca. E. mitsugo' TSYB3: GCA_964656655, 'Ca. P. opulenta' AC1: ERZ25059164.

The dsc BGC is deposited on MIBiG with the following accession number BGC0003182.

Other data related to this work (e.g. HPLC-HRMS) are available from the lead contact upon request because they might include data for ongoing projects that will be processed at a later stage of this project.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

N/A

Population characteristics

N/A

Recruitment

N/A

Ethics oversight

N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

The sample size (number) of analysed sponges was determined by the occurrence of 'Ca. Entotheonella' phylotypes and the availability of sponge samples. Sampling sponges is not an easy and cheap thing to do. This is why we had usually only one sponge sample per specimen.

Data exclusions

No data were excluded from this study.

Replication

All protein expressions were repeated multiple times, with consistent outcomes that were measured. Enzyme assays were performed with freshly purified enzyme and in triplicates and were all successful and had the same outcome.

Randomization

No randomization of samples was necessary to this study because we looked at specific samples and wanted to link sponge metabolites to their respective genomic origin. We needed to know what the respective sample contains.

Blinding

Blinding was not necessary to this study, because we looked at the specific samples and wanted to link sponge metabolites to their respective genomic origin. We needed to know what the respective sample contains.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging