Main Manuscript for

Single-cell metabolite detection and genomics reveals uncultivated talented producer

Masato Kogawa^{1,2*}, Rimi Miyaoka^{1*}, Franziska Hemmerling^{3*}, Masahiro Ando⁴, Kei Yura^{1,4,5}, Keigo Ide^{1,2}, Yohei Nishikawa^{1,2}, Masahito Hosokawa^{1,4}, Yuji Ise⁷, Jackson K. B. Cahn³, Kentaro Takada⁸, Shigeki Matsunaga⁹, Tetsushi Mori¹⁰, Jörn Piel³ & Haruko Takeyama^{1,2,4,6}

¹ Department of Life Science and Medical Bioscience, Waseda University, 2-2 Wakamatsu-cho, Shinjuku-ku, Tokyo, 162–8480, Japan

² Computational Bio Big-Data Open Innovation Laboratory, AIST-Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo, 169–0072, Japan

³ Institute of Microbiology, Eidgenössische Technische Hochschule (ETH) Zurich, Vladimir-Prelog-Weg 4, 8093 Zurich, Switzerland

⁴ Research Organization for Nano & Life Innovation, Waseda University, 513 Wasedatsurumaki-cho, Shinjuku-ku, Tokyo, 162–0041, Japan

⁵ Graduate School of Humanities and Sciences, Ochanomizu University, 2-1-1 Otsuka, Bunkyo-ku, Tokyo, 112-8610, Japan

⁶ Institute for Advanced Research of Biosystem Dynamics, Waseda Research Institute for Science and Engineering, Graduate School of Advanced Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo, 169-8555, Japan

⁷ Sesoko Station, Tropical Biosphere Research Center, University of the Ryukyus, 3422 Sesoko, Motobu, Kunigami, Okinawa, 905-0227, Japan

⁸ School of Marine Biosciences, Kitasato University, Minami-ku, Sagamihara, Kanagawa, 252-0373, Japan

⁹ Laboratory of Aquatic Natural Products Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, 113-8657, Japan

¹⁰Department of Biotechnology and Life Science, Tokyo University of Agriculture and Technology, Koganei, Japan

* Equally contributed

✉ Correspondence

Correspondence:

Haruko Takeyama (haruko-takeyama@waseda.jp)

Jörn Piel (jpiel@micro.biol.ethz.ch)

Author Contributions: H.T., J.P., M.H., M.A., T.M. conceived the study. S.M., K.T. and Y.I. collected the sponges and determined the chemotypes, S.M. and K.T. conducted MS analysis of the sponges. R.M. and M.A. performed Raman microspectroscopy, M.K. and Y.N. conducted metagenomic and single-cell genomic studies. M.K., K.I., and K.Y. analyzed genomic data. F.H., J.K.B.C. and J.P. analyzed biosynthetic gene clusters and F.H. conducted biochemical studies on auranosides. All authors planned the experiments, and M.K., F.H., K.Y., J.P., and H.T. wrote the manuscript.

Competing Interest Statement: The authors declare no competing financial interests.

Classification: BIOLOGICAL SCIENCES, Microbiology

Keywords: natural product, single-cell genomics, Raman spectroscopy

ABSTRACT

The production of bioactive metabolites is increasingly recognized as an important function of host-associated bacteria. An example is defensive symbiosis that might account for much of the chemical richness of marine invertebrates including sponges (Porifera), one of the oldest metazoans. However, as most complex microbiomes remain largely uncultivated and lack reference genomes, unequivocally linking metabolic functions to a cellular source is a challenge. Here we report an analysis pipeline of microfluidic encapsulation, Raman microscopy, and integrated digital genomics (MERMAID) for an efficient identification of uncultivated producers. We applied this method to the chemically rich bacteriosponge *Theonella swinhoei*, previously shown to contain 'Entotheonella' symbionts providing most of its bioactive substances except for the antifungal aurantosides that lacked biosynthetic gene candidates in the metagenome. Raman-guided single-bacterial analysis and sequencing revealed a cryptic, distinct multiproducer, '*Candidatus* Poriflexus aureus' from a new Chloroflexi lineage. Its exceptionally large genome contains numerous biosynthetic loci and suggested an even higher chemical richness of this sponge than previously appreciated. This study highlights the importance of complementary technologies to uncover microbiome functions, reveals remarkable parallels between distantly related symbionts of the same host, and adds functional support for diverse chemically prolific lineages being present in microbial dark matter.

Significance Statement

The production of bioactive metabolites is increasingly recognized as an important function of host-associated bacteria. However, the acquisition of integrated genomic and metabolic data from uncultivated environmental bacteria is still challenging. In this work, we explored the combination of Raman microscopy and single-cell sequencing to localize chemical features to a specific bacterium in an uncultivated microbiome, and we specified the bacteria in the uncultured lineage as a producer of aurantoside, an antifungal natural product, from a chemically and microbially complex sponge. This study offers a new methodology as well as insights into chemical functions of uncultivated life.

Main Text

Introduction

Sponges (Porifera) are among the oldest metazoans and thrive in a wide variety of habitats from warm tropical coral reefs to cold deep oceans and freshwater lakes and streams. Their ability to process and filter massive volumes of water during feeding casts a huge impact on biogeochemical cycling¹ and exposes sponges to a wide range of microorganisms². Many microbes escape lysis by sponges and establish microbiomes that vary from few bacteria to spectacular consortia comprising a major portion of the host biomass. With evidence for a wide range of functional roles that include the acquisition of nutrients, stabilization and reshaping of the sponge skeleton, and disposal of metabolic waste and toxic compounds^{3,4}, sponge-microbe associations might contribute to the success of their evolutionarily ancient animal hosts in multiple ways. Characterizing such roles and identifying key microbiota, however, remains a major challenge due to a high complexity of many sponge microbiomes and low success rates of cultivation⁵.

Many sponges contain rich sets of bioactive metabolites with demonstrated or proposed roles in chemical protection and colonization⁶, and as sources of new therapeutics⁷. Recent studies, enabled by metagenomic methods, have uncovered key roles of some sponge symbionts in synthesizing specialized metabolites for their hosts⁸⁻¹¹. A remarkable example is the chemically rich sponge *Theonella swinhoei* (Fig. 1a), in which a single to few 'Candidatus Entotheonella' producers within highly diverse microbiomes generate numerous bioactive compounds in their hosts^{9,12,13}. 'Entotheonella' are members of the new candidate phylum 'Tectomicrobia' (Entotheonellaeota)⁸ and form large multicellular filaments that can be enriched by mechanical fractionation¹⁴. Metagenomic and single-bacterial sequencing of such enriched microbiomes (Fig. 1a) revealed in several *T. swinhoei* chemotypes large sets of biosynthetic gene clusters in 'Entotheonella' genomes that were connected to virtually all known sponge polyketides and peptides^{8,9} (Fig. 1b). Besides suggesting important and potentially widespread contributions by sponge microbiota to host defenses, the results demonstrated that uncultivated life harbors lineages with a chemical richness comparable to that of well-known "talented" and industrially important producer taxa, such as filamentous actinomycetes¹⁵.

The identification of metabolic sources in complex microbiomes remains a formidable task. While total metagenomic sequencing can cover much of the sequence space, contigs harboring biosynthetic genes often remain unassigned due to lack of

informative taxonomic marker genes or to divergent codon usage that prevent binning. Single-cell genomics, on the other hand, can provide this information but might require the analysis of large sample numbers until a target cell is identified. Here we asked whether an analysis pipeline of microfluidic encapsulation, Raman microscopy, and integrated digital genomics (MERMAID) can accelerate the search for key chemical contributors. Tested on a case for which previous methods had revealed neither the organism nor biosynthetic genes, the process rapidly identified a cryptic producer of antifungal metabolites in *T. swinhoei*. The symbiont, a member of a new, deep-rooting *Chloroflexi* lineage that is morphologically similar to 'Entotheonella', is highly abundant in the filamentous bacterial fraction but has eluded metagenomic sequencing. On its exceptionally large 14 Mbp genome, numerous biosynthetic gene clusters (BGCs) in addition to the aurantoside locus were identified, suggesting an even higher potential for metabolic richness of this sponge than previously appreciated. Our study highlights the importance of complementary technologies to comprehensively study microbiomes, reveals remarkable parallels between distantly related symbionts of the same host, and adds functional support for diverse chemically prolific lineages present in microbial dark matter.

Results

An elusive biosynthetic source of aurantosides in *Theonella swinhoei* Y

Among the diverse characterized bioactive polyketides and peptides known from *Theonella swinhoei* Y ("Y" refers to the chemotype with yellow interior) (Fig. 1b), the aurantosides represented an exception as the only compounds that could not be assigned to 'Entotheonella'⁸. Aurantosides comprise a group of yellow-pigmented, structurally related halogenated tetramic acid glycosides with selective antifungal¹⁶ or cytotoxic¹⁷ activities. Based on the aurantoside structure, we expected the pathway to involve a polyketide synthase-nonribosomal peptide synthetase (PKS-NRPS) for the polyene-tetramic acid core structure, one or more halogenases, and enzymes for sugar biosynthesis and attachment. In the prior studies on *T. swinhoei* Y, metagenomic binning analyses showed that its mechanically enriched filamentous bacterial fraction contained two 'Entotheonella' variants, 'Entotheonella factor' and 'Entotheonella gemina', with distinct sets of BGCs. These variants appeared to correlate with the presence of two related but discriminable filament morphotypes in the bacterial preparation (Fig. 1a) that

had been assumed to represent the 'Entotheonella' variants. However, efforts failed to identify matching gene candidates for auranoside biosynthesis in the BGC-rich 'Entotheonella' genomes or the highly fragmented non-'Entotheonella' metagenomic bins.

When conducting new cell-separation studies with freshly collected *T. swinhoei* Y specimens, we noticed the presence of large yellow bacterial filaments that lost pigmentation after several washing steps with artificial sea water (ASW). In the bleached state, the bacteria appeared identical to the morphotype that we had previously assumed to be one of the two 'Entotheonella' variants present in the sponge. Since the 'Entotheonella' genomes lacked discernible candidates for an auranoside BGC but contained genes for carotenoid biosynthesis¹⁸, the identity of the pigment remained unresolved. This conundrum represented an ideal test case for an integrated Raman microscopy-single bacterial genomics approach.

The MERMAID single-cell analysis platform

The MERMAID workflow (Fig. 2) starts with single-bacterial encapsulation into microdroplets. For this purpose, a bacterial suspension is introduced into a previously fabricated microfluidic device¹⁹, and droplets containing single bacteria are generated by dispersion in an oil phase. Then, droplets are isolated into multi-well plates with small contamination risk, and Raman spectra of single bacteria are measured in each well. Finally, bacteria that show Raman bands matching to target metabolites are selected and subjected to single-cell genome amplification in the well. By next-generation sequencing of the amplified genome, selective genome analysis of bacteria containing characteristic metabolites can be achieved.

Raman microscopy-based identification of an auranoside producer candidate

To identify the auranoside source in the microbiome of *T. swinhoei* Y, we initiated this study by acquiring a Raman spectrum from a standard sample of auranoside A, a chlorinated polyene tetramic acid containing D-xylopyranose, D-arabinopyranose, and C2-methylated 5-deoxyarabinofuranose residues¹⁷ (Fig. 3a). The spectrum showed characteristic bands at 1,547 cm⁻¹ and 1,149 cm⁻¹ corresponding to the C=C stretch and the C-C stretch in the hydrophobic polyene moiety.

For single-filament Raman analyses, *T. swinhoei* Y-associated filamentous bacteria were segregated from the sponge tissue by centrifugation, and Raman mapping measurements of single filaments were conducted using the bacterial suspension directly without prior sorting. Subsequently, by applying a multivariate analysis, namely multivariate curve resolution-alternating least squares (MCR-ALS)²⁰, to all the spectral data obtained from all the measured filaments, we obtained information on several molecular components. Fig. 3b shows the MCR-ALS-decomposed spectra. Component 1 shows the characteristic bands at 1,546 and 1,149 cm⁻¹, which can be assigned to aurochromes. Component 2 shows a similar spectral feature as component 1, but the peaks have shifted to 1,600 and 1,167 cm⁻¹. Polyenes show peak shifts of C=C stretch and C-C stretch bands, depending on the length of the conjugated double bonds^{21,22}. Therefore, component 2 is assigned to a molecule with shorter polyene length inside the filaments. On the other hand, component 3 is assigned to a carotenoid based on 1511, 1155, and 1005 cm⁻¹ bands that correspond to C=C stretch, C-C stretch, and methyl rock, respectively. Among these bands, the band at 1005 cm⁻¹ was observed only in component 3, but not in components 1 and 2, consistent with carotenoids carrying methyl groups in the conjugated chain^{21,22} in contrast to aurochromes. Components 4 and 5 are assigned to common proteins and cytochromes B and C, as discussed in our previous study²³. The MCR-ALS-decomposed Raman images revealed a subset of filamentous bacteria as strong candidates for aurochrome production. As shown in Fig. 3c, aurochromes were only detected in the yellow-colored filaments (columns A and B), but not in a second subset of almost colorless filaments (column C). On the other hand, this colorless phenotype exclusively generated the carotenoid-type spectrum of component 3. Proteins and cytochromes were distributed throughout the cells in all the filaments.

With MCR-ALS-extracted spectra as references, we next tested whether it is possible to establish a microbial characterization pipeline that combines Raman microscopy with single-bacterial sequencing and thus directly links metabolomes to genomes. Raman spectra of droplet-encapsulated single filaments isolated in 384 well plates were individually obtained and analyzed as the linear combination of reference spectra. With an analysis time of a few minutes per well including focusing the cell and 1-second laser irradiation, the method rapidly provided multiple single-bacterial filaments

with aurantoxide spectral features, thus enabling efficient single-bacterial screening based on metabolite content.

Single-filament sequencing and genome assembly reveals a cryptic member of the *T. swinhoei* Y microbiome

All aurantoxide-positive bacteria identified by Raman microscopy exhibited a uniform morphology consisting of large multicellular filaments comparable in shape to those of 'Entotheonella'. Characteristic features distinct from 'Entotheonella' include a sheath-like structure enclosing each organism and a larger size. To further characterize these bacteria, genomes of 12 filaments with matching Raman spectra were amplified and individually sequenced by Illumina MiSeq, generating a total of 5.25 Gb data at medium quality ($\geq 50\%$ completeness, $< 10\%$ contamination, Table S1). To construct a high-quality draft genome, the sequence data from the 12 filaments were co-assembled by SPAdes and ccSAG, and the acquired genome was evaluated by QUAST, CheckM, and Prokka²⁴⁻²⁸ (Table S2). The first draft consisted of 3,409 contigs with an overall GC content of 49.7% and was estimated to show 95.6% completeness and 6.2% redundancy based on single-copy marker gene analysis. One rRNA operon and 41 tRNA genes were recovered (Table S3). A BLAST analysis of the 16S rRNA gene sequence retrieved a distant homolog (88.9% identity) from the *Chloroflexi* member *Caldilinea aerophila* (NR_074397) as the closest relative registered in the NCBI 16S RefSeq nucleotide database. When reanalyzing previous metagenomic datasets of the filamentous preparation⁸, we found that the genome of this organism was poorly covered by short contigs despite its abundance and had therefore remained unnoticed. The bacterium was named '*Candidatus Poriflexus aureus*' TSY due to its association with Porifera and its yellow color.

With 14 Mbp, the calculated size of the draft genome exceeded even the large 9-10 Mbp 'Entotheonella' genomes and is one of the largest in the *Chloroflexi* genomes (Table S4), but it is definitely the largest among the recently reported 1,200 sponge symbionts²⁹. Since this piece of data appeared highly unusual, we further verified the genome quality by the following investigations (Fig. S1): Individual size estimations conducted from each single-amplified genome (SAG), binning of the contigs based on sequence depths using newly acquired metagenomic shotgun sequencing data, and taxonomic annotation of each contig. Metagenomic sequencing was accomplished with

a modified, gentler bacterial lysis protocol after noticing that the *Chloroflexi* filaments rapidly disintegrated during DNA preparation in comparison to 'Entotheonella'. The size and high quality of the draft genome was confirmed by these analyses, making this one of the largest bacterial genomes reported to date. To further improve the assembly and generate longer contigs for BGC studies, long-read sequencing of SAG mixture was conducted using MinION nanopore technology. By hybrid assembly using long reads and SAG short reads, 195 contigs including one plasmid with a contig N50 of 102,784 bp were acquired (Table S2, Fig. S2). In summary, the data obtained by Raman microscopy and construction of multiple SAGs of 'P. aureus' TSY provide evidence for a cryptic filamentous aurantoside producer with a large 14 Mbp draft genome and at least one plasmid.

To test whether 'P. aureus' phylotypes are present in other *Theonella* sponges, we investigated a blue specimen of unknown species affiliation containing aurantosides that had been collected around Shimoji island 1,700 km away from Hachijo-jima island (Fig. S3), the collection site of *T. swinhoei* Y. In agreement with our previous data, cell dissociation experiments revealed sheathed filaments that looked identical to those of *T. swinhoei* (Fig. S3). Raman spectra obtained from isolated filaments of the blue sponge were analyzed by MCR-ALS analysis (Fig. S3). In addition to a congener with an aurantoside A-type chromophore, MCR-ALS decomposition suggested the presence of a variant with shortened polyene system based on band shifts to higher frequency. A draft genome of this symbiont was acquired with 68.2% completeness (Table S2). The nucleotide identity of its 16S rRNA gene to that of 'P. aureus' TSY was more than 99%. Likewise, the average nucleotide identity of the two draft genomes was calculated as 98.0%, suggesting that both organisms belong to the same candidate species. Because the estimated genome size of the newly identified symbiont was 12.9 Mbp, the large genome was further supported. The symbiont from the blue *Theonella* sp. will be referred to as 'P. aureus' BT hereafter.

Identification and functional analysis of the aurantoside biosynthetic gene cluster

Based on our biosynthetic hypothesis for aurantosides, we searched the 'P. aureus' TSY genome for matching gene candidates. This analysis retrieved two small BGC fragments that each contained one halogenase gene, a more extended contig encoding a portion of a trimodular PKS-NRPS, and a contig harboring genes for a

putative S-adenosylmethionine-(SAM)-dependent methyltransferase and for sugar biosynthesis and attachment. Combinatorial PCR experiments connected these fragments, confirming that they belong to the same locus. Further extension was achieved by MinION sequencing, resulting in a candidate aurantoside (*ats*) BGC with 17 ORFs and of 38 kbp length (Fig. 4a, Table S5).

In order to functionally test the involvement of the *ats* BGC in aurantoside production, we first attempted to perform heterologous expression experiments on the PKS protein AtsF with or without the halogenases AtsD or AtsE. Since none of these trials yielded soluble proteins, we next focused on the methyltransferase homolog AtsJ sharing the highest similarity to the members of the FkbM family (Table S5) that catalyze various O-methylations, such as in the polyketide pathways for the immunosuppressant FK506³⁰ and the lobophorins³¹. As the only methyltransferase encoded in the BGC, AtsJ represented a strong candidate to convert aurantoside B (**1**) into its O-methylated congener aurantoside A (**2**) (Fig. 4b). Initial screens were conducted in cell-free assays with recombinant AtsJ and an authentic standard of **1** isolated from *T. swinhoei* Y. Subsequent assays with purified enzyme contained AtsJ in assay buffer, SAM, and **1**. We observed the conversion of **1** into **2** only in the presence of AtsJ (Fig. 4c), supporting a role of the identified BGC in aurantoside biosynthesis. Controls without SAM, without AtsJ, or with boiled enzyme gave no conversion, while controls without the cofactor Mg²⁺ gave partial conversions, presumably due to the presence of low Mg²⁺ concentrations in the buffer or enzyme preparation (Fig. S4).

Based on these experimental results, we propose a biosynthetic pathway for aurantoside formation (Fig. 4b). A polyene chain is assembled from six malonyl building blocks in an iterative fashion by AtsF and chlorinated twice by AtsD and AtsE. AtsG catalyzes further extension steps with a malonyl and an Asn unit, before the thioesterase domain releases the aurantoside aglycone as a tetramic acid by Dieckmann cyclization. Subsequently, three glycosidic units are modified by AtsK, AtsL, and AtsN, and transferred by the glycosyl transferase homologs AtsH, AtsI, and AtsM. In a last step, AtsJ catalyzes the methylation of **1**, yielding congener **2**.

The '*P. aureus*' genome contains a large number of orphan BGCs

While searching the '*P. aureus*' TSY genome for the aurantoside BGC, we unexpectedly detected numerous additional contigs with natural product genes

suggested by the bioinformatics tool antiSMASH (version 5.0)³² and manual analysis (Fig. 4d). Based on the hybrid assembly draft, extensive gap closing experiments by combinatorial PCR and metagenomic cosmid isolation and sequencing generated a high-quality dataset of 19 additional extended biosynthetic loci that belong to diverse natural product classes (Fig. 4d). Omitted from this list are BGCs containing less than three biosynthetic domains and a phytoene synthase-encoding locus identified by antiSMASH that likely belongs to primary metabolism. All BGCs were localized in the 'P. aureus' core genome, except for the BGC *pau3* encoding a bimodular NRPS, which mapped to the single-contig plasmid (Fig. S2). In addition to the conventional BGC *pau1* representing a nine-module hybrid type I *cis*-AT PKS-NRPS, a large number of architecturally unusual *trans*-AT PKS and NRPS BGCs and BGC fragments were identified (*pau2-pau14*). Further BGCs belong to terpene (*pau15*) and ribosomally biosynthesized and posttranslationally modified peptide (RiPP) pathways (*pau16-pau19*). In addition to structural predictions based on tools integrated into antiSMASH, we used TransATor³³, a recently developed bioinformatics tool for the automated prediction of polyketide structures for *trans*-AT PKSs. All predictions for substrate specificities for PKS and NRPS domains are annotated in Fig. 4d, but these analyses did not suggest a known natural product for any of the BGCs. Similarly, KnownClusterBlast analysis showed low similarities of the 'P. aureus' BGCs to characterized clusters, suggesting new natural product skeletons. To our knowledge, a comparable BGC richness among uncultivated bacteria has to date only been reported from 'Entotheonella', further adding to the remarkable similarities between these unrelated microbes inhabiting the same sponge.

'Poriflexia', a novel lineage of *Porifera*-associated filamentous *Chloroflexi*

In addition to their unusual genome size greatly exceeding that of previously reported *Chloroflexi*, 'P. aureus' TSY exhibits, to the best of our knowledge, a new morphology for this phylum. Most of the filamentous *Chloroflexi*, which belong to the classes *Chloroflexia*, *Anaerolineae*, and *Caldilinea*, are highly elongated, being longer than 100 μm and <1-2 μm in width, and their filaments bend flexibly. In contrast, 'P. aureus' TSY filaments are 30-100 μm in length and 5-10 μm wide, covered by a sheath, and straight. To further assess their affiliation, a phylogenetic tree based on 43 bacterial marker genes was constructed using genomes of filamentous *Chloroflexi* in the Refseq

database (Fig. 5a). Concatenated sequences of 6,988 amino acid residues were generated from translated marker genes and aligned by CheckM. While most known filamentous *Chloroflexi* bacteria belong to the classes *Chloroflexia* or *Anaerolineae*, the phylogram supported a distinct, deep-branching clade for '*P. aureus*'. This phylogenetic dataset also suggested that members of the class *Caldilineae* are most closely related to '*P. aureus*', in agreement with the 16S rRNA gene-based phylogenetic analysis (Fig. S5). For the new deep-rooting *Chloroflexi* lineage we propose the name 'Poriflexia', which seems to be different from the known Porifera-associated *Chloroflexi* (Fig. S6)²⁹.

Metabolic pathway of '*P. aureus*'

In the draft genome of '*P. aureus*' TSY, gene sets for glycolysis (Embden-Meyerhof-Parnas), tricarboxylic acid (TCA) cycle, the pentose phosphate pathway and a respiratory chain were found (Fig. 5b), suggesting that '*P. aureus*' is heterotrophic. Based on analyzing preserved metabolic pathways, the symbiont seems to perform the assimilatory reduction of sulfate and lack the nitrate/nitrite reduction pathway, features resembling those of other sponge-associated *Caldilineae* bacteria³⁴. '*P. aureus*' is prototroph for vitamins B₃ and B₆ and auxotroph for most B-vitamins (B₁, B₂, B₅, B₇, B₉ and B₁₂), similar to its closest characterized homolog, *Caldilinea aerophila* DSM 14535 which is prototroph for B₆ and B₁₂ and auxotroph for all other B-vitamins³⁵ (Table S6). For thiamine (B₁-pathway), riboflavin (B₂) and cobalamin (B₁₂) uptake, homologs to known transporters from *Chloroflexus aurantiacus* are encoded in the '*P. aureus*' genome³⁵ (Table S7). In addition, known virulence mechanisms like type III or type VI secretion systems were not found.

An orthologue analysis using Orthofinder³⁶ detected 518 core orthologous gene groups across the dataset of published filamentous *Chloroflexi* genomes (Table S4). The total number of core CDSs is 2.2 times higher in the '*P. aureus*' TSY genome than in the other *Chloroflexi* genomes (Fig. S7). The symbiont genome contains highly amplified CDSs annotated in the COG database as "amino acid transport and metabolism", "carbohydrate transport and metabolism", "inorganic ion transport and metabolism", and, as highlighted above, "secondary metabolites biosynthesis, transport and catabolism". The amplification ratios were from 3.29 to 6.20 (Fig. S7), suggesting a strong bias to transport and metabolism function. The gene enrichment pattern of '*P. aureus*' BT resembles that of '*P. aureus*' TSY. Further investigation of protein functions by

InterProScan showed that 8,140 CDSs in the orthologous gene groups of 'P. aureus' TSY are associated with 2,229 Pfam families (Table S8). Of these, 158 families were enriched in the symbiont (p-value < 0.01) compared with other *Chloroflexi* genomes, and the enrichment patterns were more similar to other sponge symbionts rather than *Chloroflexi* spp. (Fig. S8). Among the protein families, an enrichment in 775 ATP-binding cassette (ABC) transporter proteins is striking. The functional classification of ABC transporter proteins was visualized by SimGraph, which clustered the putative orthologous CDSs. About 50% of ABC transporter proteins were found to form ten large clusters (Fig. 6a). The sizes of the top five clusters were extremely large compared with SimGraph plots on other bacteria (Fig. 6b). Most of the transporter proteins were annotated as components of peptide uptake (clusters 1, 2, 4, 9, and 10), sugar uptake (clusters 3 and 5), or ligand export (clusters 6, 7, and 8) systems. This transporter enrichment pattern might be a consequence of a symbiotic lifestyle involving the acquisition of host metabolites and export of bioactive natural products. Note that the peptide importer clusters contain many glutathione importers (GsiA, C, D) although the glutathione biosynthetic pathway was detected only in the 'P. aureus' TSY genome among the filamentous *Chloroflexi* genomes. These facts suggest that glutathione should have a yet-to-be-uncovered important role in the 'Poriflexus' survival. The highly duplicated ligand multidrug transport systems (YheI, MsbA) might defend 'Poriflexus' from various secondary metabolites including those derived from the 'Entotheonella' variants in the sponge holobiont, in addition to exporting its own bioactive substances.

Discussion

In this study, we present the MERMAID analysis pipeline that combines Raman microscopy-based screening with single-cell analysis to link metabolic with genomic features in uncultivated bacteria. Raman microscopy is one of the few single-cell techniques that provides a direct link between metabolites and source organism in a non-destructive way and at the single-cell level and can be applied to diverse questions in microbiological research using commercially available instrumentation. Raman microscopy was previously applied to single-bacterial natural product imaging³⁷ and to bacterial sorting based on deuterium incorporation³⁸, but to our knowledge MERMAID is the first single-cell pipeline that also provides genomic information. This pipeline can be a one-stop, high-throughput system based on microfluidics. Recently, a single-cell

analysis method linking metabolic and genomic features has been developed based on labeling of synth(et)ase enzymes³⁹. The potential advantage of MERMAID is the detection of various types of intracellular biomolecules with specific molecular features. Some Raman-active moieties that do not overlap with those of basic metabolites are conjugated olefins, alkynes, carbon-halogen bonds, heteroaromatic rings, i.e., features present in many secondary metabolites. In some cases, compounds with more common features can also be targeted using analysis with MCR-ALS³⁷. As an example for its utility, the platform enabled the rapid identification of the unusual symbiont 'P. aureus', a key provider of bioactive substances in microbiomes of two different *Theonella* sponges. The organism had remained obscure due to its morphological resemblance to 'Entotheonella' and low genome coverage in previous metagenomic sequencing rounds, emphasizing the importance of complementary technologies to obtain insights into microbiome functions.

'P. aureus', a member of a new deep-branching *Chloroflexi* lineage, shares several noteworthy features with the distantly related 'Entotheonella', including a large genome size, filamentous morphology, and a high number of natural product BGCs suggesting a rich specialized metabolism. BGC richness is relatively uncommon among bacterial lineages and has been reported for only a few organisms from microbial dark matter^{8,40,41}. The genome sizes of 'P. aureus' and 'Entotheonella' members exceed those of all other previously reported >1200 sponge-associated bacteria²⁹ by at least 3 Mb in size. The remarkable presence of two distinct non-classical producers with expanded genomes and specialized metabolism in the same sponge might be due to selective mechanisms that foster interactions with chemically rich bacteria. Among the benefits that a diversified chemistry could provide to the sessile animals are host defense against predators and epibionts, killing of eukaryotic prey, and in the case of auranosides, protection against fungal pathogens.

Chloroflexi are one of the most common phyla present in sponge microbiomes³⁴. The phylum includes metabolically highly diverse members such as heterotrophs and non-oxygenic photosynthetic autotrophs, aerobes and anaerobes, thermophiles, halophiles, and dehalogenating respirators. Since the cultivated representatives cover a comparably low phylogenetic diversity of *Chloroflexi* and the accumulated genomic information is insufficient, many gene functions in the 'P. aureus' draft genomes are still unknown. On the other hand, genome-wide analysis detected more BGC candidates than expected

besides the auranoside cluster. This high chemical potential of 'P. aureus' is similar to 'Entotheonella' and suggests a hidden natural product diversity that is much larger than the already remarkably rich described chemistry of the *T. swinhoei* holobiont. Moreover, gene enrichment analysis of the 'P. aureus' TSY genome revealed numerous ABC transporter genes and other functional genes commonly enriched in uncultured sponge-associated bacteria like 'Entotheonella' or 'Poribacteria', suggesting further functional commonalities between these symbionts.

Materials and Methods

Sponge collection

Specimens of the marine sponge *Theonella swinhoei* Y ("Y" refers to the chemotype with yellow interior containing onnamides, polytheonamides, and auranosides) were collected at a depth of 15 m into a bucket filled with seawater around Hachijo-jima island (33°8'16"N, 139°44'5"E) in Tokyo, Japan in November 2017. The blue *Theonella* sp. was collected at a depth of 12 m from the west side of Shimoji island (24°49'24"N, 125°08'07"E) in Okinawa, Japan in October 2018. The collected sponges were cut into pieces and stored at 4 °C in Ca- and Mg-free artificial sea water (CMF-ASW)⁴².

Preparation of bacterial fractions

To prepare filamentous bacterial fractions, an adjusted protocol based on a previously reported method⁸ was used. A 1 cm-square block of sponge tissue stored in CMF-ASW was minced in a 25 ml tube (IWAKI) by a surgical knife and incubated for 15 minutes in 10 ml of CMF-ASW on ice. During the incubation, the mixture was stirred every 5 minutes to suspend bacteria. Then, the mixture was passed through a 40 µm mesh, and the flow-through was collected in a 15 ml tube. The collected flow-through was centrifuged at 1,000 ×g for 10 minutes to pellet filamentous bacteria. Supernatants including unicellular organisms were removed. After resuspension of the pellet in 5 ml CMF-ASW, steps from the centrifugation to the resuspension were repeated several times. The mixture was allowed to settle for 30 seconds and supernatants were transferred to a clean 15 ml tube twice to remove dense sponge tissue. The enriched filamentous bacterial fraction was pelleted again and resuspended in 2 ml of CMF-ASW.

Isolation of single filamentous bacteria

The storage medium of the filamentous bacterial fraction was replaced by Dulbecco's Phosphate Buffered Saline liquid (DPBS, Thermo Fisher Scientific) with 1.5% ultra-low gelling temperature agarose A5030 (Sigma-Aldrich). To reduce the risk of contamination introduced by manual picking of single filaments, bacteria were encapsulated into microdroplets at a concentration of 0.1 particle/droplet by a previously fabricated microfluidic device¹⁹. The cell suspension with 1.5×10^3 particles/ μL and 2% Pico-Surf™ 1 in Novec™ 7500 (Dolomite) as carrier oil were loaded to the microchannel through polytetrafluoroethylene (PTFE) tubing (AWG24) by a Mitos P-pump (Dolomite), and droplets (50 μm in diameter) were generated. Collected droplets were solidified on ice and moved into the water phase from the oil phase as previously reported⁴³. Droplets containing single filaments were manually picked using a micropipette (Drumond) under microscopic observation and isolated into a 384-well glass bottom (Corning) plate with 1.5 μL of DPBS.

Raman spectroscopy

All Raman spectroscopic measurements were performed by using a laboratory-built 532-nm-excited Raman microspectrometer, as described in our previous report²³. With a confocal setup, by a 100 $\times$, 1.4-NA objective lens both for laser excitation and scattered light collection, the spatial resolution of Raman imaging system was $0.3 \times 0.3 \mu\text{m}$ in the lateral directions and $2.6 \mu\text{m}$ in the axial direction. The spectral resolution was 3.0 cm^{-1} . For measuring the standard spectra, a solution of aurantioside A in DMSO was prepared. The solution was placed between two coverslips, and sealed with nail polish. The laser power and exposure time were set to 15 mW and 60 seconds, respectively. For the Raman imaging measurements of filamentous bacteria, 1 μL bacterial suspension was also placed between two clean coverslips, and sealed with Vaseline. By controlling the amount of suspension medium, it was possible to physically immobilize the bacteria. In the mapping measurements, the sample stage was automatically translated with $0.5 \mu\text{m}$ interval, and the exposure time was 1 second for each measured point. For the screening of aurantioside-producing bacteria, each filament bacterium isolated in 384-well plate was measured at single-cell resolution with 1 second exposure time. In the measurement of bacteria, the laser power was set 0.5 – 5 mW.

In the data preprocessing, the obtained Raman spectral data were wavenumber-calibrated, intensity-corrected, and noise-reduced by singular value decomposition analysis (SVD), using IGOR Pro software (WaveMetrics, Inc., Lake Oswego, OR, USA). Following the data preprocessing, a multivariate analysis, namely multivariate curve resolution-alternating least squares (MCR-ALS) was conducted based on the method developed in our previous study²⁰. In this study, all the spectral data obtained from bacterial mapping measurements were combined into one matrix, and the non-negative matrix factorization was performed with the SVD initialization method and L1-norm regularized ALS optimization method. All the MCR-ALS calculations were performed by an in-house program written in python, using the SciPy library (<https://github.com/mshrAndo/PyMCR>). For the screening of auranoside producing bacteria, the MCR-ALS decomposed spectra were used as reference spectra, and the presence of auranosides was detected by least-squares spectral fitting.

Whole-genome amplification of single filaments

The Raman microscopy-selected bacteria were lysed by incubation at 37 °C for 30 minutes with Ready-lyse lysozyme (Epicentre) of 10 U/μl and heat treatment at 95 °C for 10 minutes. To acquire single-amplified genomes (SAG), multiple displacement amplification (MDA) reaction mix (REPLI-g Single Cell Kit, QIAGEN) was added to the lysate and incubated at 30 °C for 3 hours, and the amplification was terminated by incubation at 65 °C for 3 minutes. As compared to the manufacturer's protocol, the volume of the MDA reactions was scaled down to 10 μl to amplify genomes in the 384-well plate. MDA products were assessed by PCR and sequencing of 16S rRNA gene fragments (V3-V4 region).

Whole-genome sequencing

Illumina sequencing libraries were prepared from each MDA product mixture using the Nextera XT kit. Libraries were sequenced in a 2 × 301-bp paired-end run by MiSeq (Illumina), and 12 SAG data of 'P. aureus' with a total size of 5.26-Gb. To evaluate the acquired draft genomes, metagenomic shotgun sequencing was also conducted using a procedure carefully adapted to lyse 'Entotheonella' filaments to enrich 'P. aureus' genomic DNA. Metagenomic DNA was extracted by DNA Clean & Concentrator Kit (ZYMO research) from lysate of the filamentous bacterial fraction acquired by

incubation for 8 hours with 10 U/μl Ready-lyse lysozyme (Epicentre). The sequencing library was prepared using the Nextera XT kit and sequenced the same way as for the SAGs.

Metagenomic cosmid library screening

To isolate cosmids harboring PKS genes, a metagenomic cosmid library, previously⁴⁴ prepared from *T. swinhoei* Y total DNA in the pWEB vector (Epicentre), was screened and a 3D growth-and-dilution library format as reported earlier⁴⁵. The isolated cosmids pTSTA2 and pTSTA3 were sequenced using Sanger technology.

Biochemical assays with the aurantoside methyltransferase AtsJ

Based on the *atsJ* sequence, a codon-optimized gene for heterologous gene expression in *E. coli* was synthesized (Integrated DNA Technologies). Flanking restriction sites for *NdeI* and *EcoRI* were added, resulting in the plasmid pUCIDT_*atsJ*. The *NdeI* and *EcoRI*-excised insert of pUCIDT_*atsJ* was ligated into the same restriction sites of pET-28b (Novagen) and the resulting plasmid was introduced into *E. coli* DH5 α by electroporation. The construct was confirmed by DNA sequencing.

For protein expression, electrocompetent *E. coli* TunerTM (DE3) cells (Novagen) were co-transformed with the plasmids pET-28b_*atsJ* and pG-KJE8 (TAKARA). A 200 ml culture in TB medium (50 μg/ml kanamycin, 20 μg/ml chloramphenicol, 0.5 mg/ml L-arabinose, and 5 ng/ml tetracycline) was grown at 37 °C and 180 rpm to an OD₆₀₀ of 0.5 - 0.8, cooled on ice for 5 minutes, induced with isopropyl β-D-1-thiogalactopyranoside (0.1 mM final concentration), then shaken for 20 hours (180 rpm, 16 °C). Cells were collected by centrifugation and homogenized in 1 mL per 100 mg dry cell material in assay buffer (AB, 50 mM phosphate buffer, pH 8.0, 300 mM NaCl, 10% [v/v] glycerol). The suspension was placed on ice, and cells were lysed by sonication. Cell debris was removed by centrifugation at 12,000 ×g at 4 °C for 20 minutes and the cleared supernatant was either directly used for the reconstitution assays or subjected to purification with Ni-NTA agarose (Macherey-Nagel). The suspension was incubated for 60 minutes and transferred to a fretted column. The resin was washed with AB with 20 mM imidazole and eluted with 3×0.5 ml elution buffer (AB with 250 mM imidazole). Subsequently, imidazole was removed using a PD-minitrapp column (GE Healthcare) for buffer exchange to AB.

The assays were conducted in a total volume of 200 μ l of AtsJ in AB, supplemented with 5 mM $MgCl_2$, 1 mM SAM and 0.25 mM aurantioside B. Negative controls with concentrated lysate from *E. coli* TunerTM (DE3) pG-KJE8 pET28b(+), boiled concentrated lysate of *E. coli* TunerTM (DE3) pG-KJE8 pET28b(+)_atsJ (10 min at 98 °C) and without addition of SAM were analyzed in parallel. After incubation for 3 hours at 30 °C and 180 rpm in the dark, samples were prepared for ultra-high performance liquid chromatography-high resolution heated electrospray-mass spectrometry (UPLC HR HESI MS) analysis in the dark using ZipTip[®]C18 (Millipore) purification and resuspension in acetonitrile.

The UPLC HR HESI MS experiments were performed on a Dionex Ultimate 3000 UHPLC coupled to a Thermo Scientific Q ExactiveTM mass spectrometer. A solvent gradient (A = water + 0.1% formic acid and B = acetonitrile + 0.1% formic acid with 95:5 A/B ramped to 100% B over 10 min, 100% B for 5 min, ramped to 95:5 A/B over 1 min at a flow rate of 1.0 ml/min) was used on a Kinetex C18-XB (2.6 μ , 150 $\times$ 4.6 mm) column at 27 °C. The MS was operated in positive ionization mode at a scan range of 100 to 1,000 *m/z* with a scan resolution of 70,000. The spray voltage was set to 3.7 kV and the capillary temperature to 320 °C. Spectra were annotated using the software XcaliburTM version 4.1 (Thermo Fisher Scientific).

Assembly and annotation of the 'Poriflexus aureus' genome

The quality of each SAG dataset was controlled by removing low-quality reads ($\geq 50\%$ of bases with quality scores ≤ 25), trimming the 3'-end low-quality bases (quality score ≤ 20), and removing < 20 bp reads and reads including $> 1\%$ unidentified bases. Contigs were assembled from each SAG dataset by SPAdes 3.12.0²⁴ with “--sc --careful” option. Then, a composite draft genome was generated by ccSAG²⁵. The assembly quality of composite draft genomes and drafts from each SAG data were evaluated by QUAST 4.5²⁶ and genome completeness and contamination rate of them were assessed by CheckM 1.0.6²⁷.

Gene annotation and metabolism prediction of the composite draft genomes was conducted by Prokka 1.13²⁸, Genomaple⁴⁶, and BLAST search⁴⁷ to COG database⁴⁸. All high-quality *Chloroflexi* genomes ($> 90\%$ genome completeness, $< 10\%$ contamination rate) except for the class Dehalococcoidia composed of unicellular *Chloroflexi* sp. and genomes of five sponge-associated bacteria (two 'Entotheonella' and three

'Poribacteria' variants) used in the comparative genome analysis were downloaded from the NCBI genome database (Table S4) and annotated by the same protocol as for the composite draft genomes. *Chloroflexi* phylogenetic trees were inferred by the maximum likelihood method using RAxML v 0.9.0⁴⁹ based on concatenated marker genes identified by CheckM or 16S rRNA genes. Functional enrichment analysis of the 'Poriflexus' genomes were conducted by Fisher's exact test using Pfam annotation result, and the result was visualized by iTOL⁵⁰. Orthofinder³⁶ was used for finding orthologous gene groups encoded in the different genome sequences and to estimate the degree of gene duplication in each genome sequence. SimGraph, an in-house tool for visualizing gene duplication based on fasta36⁵¹ and Cytoscape 3.8.0⁵², was used to compare the degree of duplication of a specific gene. Briefly, all-against-all similarity calculation was carried out among the amino acid sequences derived from the duplicated genes by fasta36 and pairs with E-value < 10⁻¹⁰ were connected with edges on Cytoscape.

Long-read sequencing and analysis of BGCs

To acquire continuous sequences of BGCs, long-read sequencing was conducted. A sequencing library was prepared by Rapid Sequencing kit (Nanopore) from a mixture of five single-filament MDA products generating high-quality SAG short reads and sequenced by MinION (Nanopore) for 48 hours. From the acquired 1.48-Gb long reads, long-read contigs were assembled by Canu 1.4⁵³ and polished by Pilon⁵⁴ using short-read sequences of 12 short-read SAG samples. Gene annotation of long-read contigs was conducted by Prokka and BGCs were identified and annotated by antiSMASH 4.0.2⁵⁵. Predictions of natural product structures for *trans*-AT PKSs were generated by TransATor³³. Manual BLAST analyses for verification of BGCs and discovering of undetected BGCs were carried out in addition.

Data availability

Sequence reads and genomic sequences of 'P. aureus' have been submitted to the National Center for Biotechnology Information Search database under project number PRJNA681468. Individual accession numbers of compared bacterial genomes can be accessed in Supplementary Table 3. The aurantoside BGC is registered to the MiBIG (ID BGC0002131).

Acknowledgments

This work was supported by MEXT KAKENHI grant numbers 17H06158 to H.T., by the Gordon and Betty Moore Foundation (grant GBMF9204, DOI: <https://doi.org/10.37807/GBMF9204>) and the Swiss National Science Foundation (205320_185077) to J.P. We thank C. Sakanashi, H. Maciejewska-Rodrigues and H. Minas for technical support. The super-computing resource was provided by the Human Genome Center (University of Tokyo).

References

1. Maldonado, M. *et al.* Sponge skeletons as an important sink of silicon in the global oceans. *Nature Geoscience* **12**, 815–822 (2019).
2. Thomas, T. *et al.* Diversity, structure and convergent evolution of the global sponge microbiome. *Nat. Commun.* **7**, 11870 (2016).
3. Mohamed, N. M., Colman, A. S., Tal, Y. & Hill, R. T. Diversity and expression of nitrogen fixation genes in bacterial symbionts of marine sponges. *Environ. Microbiol.* **10**, 2910–2921 (2008).
4. Keren, R. *et al.* Sponge-associated bacteria mineralize arsenic and barium on intracellular vesicles. *Nat. Commun.* **8**, 14393 (2017).
5. Montalvo, N. F. *et al.* Integration of culture-based and molecular analysis of a complex sponge-associated bacterial community. *PLoS One* **9**, e90517 (2014).
6. Paul, V. J., Freeman, C. J. & Agarwal, V. Chemical Ecology of Marine Sponges: New Opportunities through ‘-Omics’. *Integrative and Comparative Biology* **59**, 765–776 (2019).
7. Anjum, K. *et al.* Marine Sponges as a Drug Treasure. *Biomol. Ther.* **24**, 347–362 (2016).
8. Wilson, M. C. *et al.* An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* **506**, 58–62 (2014).
9. Mori, T. *et al.* Single-bacterial genomics validates rich and varied specialized metabolism of uncultivated *Entotheonella* sponge symbionts. *PNAS* **115**, 1718–1723 (2018).

10. 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).
11. Rust, M. *et al.* A multiproducer microbiome generates chemical diversity in the marine sponge *Mycale hentscheli*. *PNAS* **117**, 9508–9518 (2020).
12. Freeman, M. F., Vagstad, A. L. & Piel, J. Polytheonamide biosynthesis showcasing the metabolic potential of sponge-associated uncultivated ‘Entotheonella’ bacteria. *Curr. Opin. Chem. Biol.* **31**, 8–14 (2016).
13. 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’. *Marine Biology* **136**, 969–977 (2000).
14. 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).
15. Jaspars, M. & Challis, G. Microbiology: a talented genus. *Nature* **506**, 38–39 (2014).
16. Kumar, R., Subramani, R., Feussner, K.-D. & Aalbersberg, W. Aurantoside K, a new antifungal tetramic acid glycoside from a Fijian marine sponge of the genus *Melophlus*. *Mar. Drugs* **10**, 200–208 (2012).
17. Matsunaga, S., Fusetani, N., Kato, Y. & Hirota, H. Aurantosides A and B: cytotoxic tetramic acid glycosides from the marine sponge *Theonella* sp. *Journal of the American Chemical Society* **113**, 9690–9692 (1991).
18. 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. *PNAS* **114**, E347–E356 (2017).
19. Nishikawa, Y. *et al.* Monodisperse Picoliter Droplets for Low-Bias and Contamination-Free Reactions in Single-Cell Whole Genome Amplification. *PLoS One* **10**, e0138733 (2015).
20. Ando, M. & Hamaguchi, H.-O. Molecular component distribution imaging of living cells by multivariate curve resolution analysis of space-resolved Raman spectra. *J. Biomed. Opt.* **19**, 011016 (2014).

21. Rimai, L., Heyde, M. E. & Gill, D. Vibrational spectra of some carotenoids and related linear polyenes. A Raman spectroscopic study. *J. Am. Chem. Soc.* **95**, 4493–4501 (1973).
22. Tschirner, N. *et al.* Resonance Raman spectra of β -carotene in solution and in photosystems revisited: an experimental and theoretical study. *Physical Chemistry Chemical Physics* **11**, 11471 (2009).
23. Miyaoka, R. *et al.* In situ detection of antibiotic amphotericin B produced in *Streptomyces nodosus* using Raman microspectroscopy. *Mar. Drugs* **12**, 2827–2839 (2014).
24. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
25. Kogawa, M., Hosokawa, M., Nishikawa, Y., Mori, K. & Takeyama, H. Obtaining high-quality draft genomes from uncultured microbes by cleaning and co-assembly of single-cell amplified genomes. *Sci. Rep.* **8**, 2059 (2018).
26. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
27. 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).
28. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
29. Robbins, S. J. *et al.* A genomic view of the microbiome of coral reef demosponges. *ISME J.* (2021) doi:10.1038/s41396-020-00876-9.
30. Shafiee, A., Motamedi, H. & Chen, T. Enzymology of FK-506 biosynthesis. Purification and characterization of 31-O-desmethylFK-506 O:methyltransferase from *Streptomyces* sp. MA6858. *European Journal of Biochemistry* **225**, 755–764 (1994).
31. Yue, C. *et al.* Cloning and identification of the lobophorin biosynthetic gene cluster from marine *Streptomyces olivaceus* strain FXJ7.023. *Pak. J. Pharm. Sci.* **29**, 287–293 (2016).
32. Blin, K. *et al.* antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **47**, W81–W87 (2019).

33. Helfrich, E. J. N. *et al.* Automated structure prediction of trans-acyltransferase polyketide synthase products. *Nat. Chem. Biol.* **15**, 813–821 (2019).
34. Bayer, K., Jahn, M. T., Slaby, B. M., Moitinho-Silva, L. & Hentschel, U. Marine Sponges as Hot Spots: Genomic Insights and High-Resolution Visualization of an Abundant and Diverse Symbiotic Clade. *mSystems* **3**, e00150-18 (2018).
35. Rodionova, I. A. *et al.* Genomic distribution of B-vitamin auxotrophy and uptake transporters in environmental bacteria from the Chloroflexi phylum. *Environ. Microbiol. Rep.* **7**, 204–210 (2015).
36. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
37. Horii, S. *et al.* Detection of Penicillin G Produced by *Penicillium chrysogenum* with Raman Microspectroscopy and Multivariate Curve Resolution-Alternating Least-Squares Methods. *Journal of Natural Products* **83**, 3223–3229 (2020).
38. Lee, K. S. *et al.* Optofluidic Raman-activated cell sorting for targeted genome retrieval or cultivation of microbial cells with specific functions. *Nat. Protoc.* **16**, 634–676 (2021).
39. Kim, W. E. *et al.* Synthase-Selective Exploration of a Tunicate Microbiome by Activity-Guided Single-Cell Genomics. *ACS Chem. Biol.* **16**, 813–819 (2021).
40. Crits-Christoph, A., Diamond, S., Butterfield, C. N., Thomas, B. C. & Banfield, J. F. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. *Nature* **558**, 440–444 (2018).
41. Sharrar, A. M. *et al.* Bacterial Secondary Metabolite Biosynthetic Potential in Soil Varies with Phylum, Depth, and Vegetation Type. *MBio* **11**, (2020).
42. Spiegel, M. & Rubinstein, N. A. Synthesis of RNA by dissociated cells of the sea urchin embryo. *Exp. Cell Res.* **70**, 423–430 (1972).
43. Chijiwa, R. *et al.* Single-cell genomics of uncultured bacteria reveals dietary fiber responders in the mouse gut microbiota. *Microbiome* **8**, 5 (2020).
44. Piel, J. *et al.* Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *PNAS* **101**, 16222–16227 (2004).
45. Hrvatin, S. & Piel, J. Rapid isolation of rare clones from highly complex DNA libraries by PCR analysis of liquid gel pools. *J. Microbiol. Methods* **68**, 434–436 (2007).

46. Arai, W. *et al.* MAPLE 2.3.0: an improved system for evaluating the functions of genomes and metagenomes. *Biosci. Biotechnol. Biochem.* **82**, 1515–1517 (2018).
47. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
48. Galperin, M. Y., Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* **43**, D261–9 (2015).
49. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).
50. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
51. Pearson, W. R. Finding Protein and Nucleotide Similarities with FASTA. *Current Protocols in Bioinformatics* **53**, 3.9.1–3.9.25 (2016).
52. Shannon, P. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* **13**, 2498–2504 (2003).
53. Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research* **27**, 722–736 (2017).
54. Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).
55. Blin, K. *et al.* antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **45**, W36–W41 (2017).

Figures

Figure. 1: Natural products and filamentous bacterial symbionts of the marine sponge *Theonella swinhoei* TSY. **a)** *T. swinhoei* TSY (top) and a light micrograph of its enriched filamentous symbiont fraction containing two different morphotypes M1 and M2 (bottom). **b)** Structures of secondary metabolites extracted from *T. swinhoei* TSY. Auranosides are shown in blue.

Figure. 2: MERMAID workflow. The droplets containing single bacterial filaments are isolated into the well, and Raman spectra of encapsulated bacteria are measured. After droplet disruption by plate shake, cell lysis and genome amplification are implemented in the well.

Figure 3: **a)** Raman spectrum of an aurantioside A standard. **b)** MCR-ALS decomposed spectra of bacterial filaments associated with *T. swinhoei* Y. **c)** Phase contrast and MCR-ALS-decomposed images of bacterial filaments (column 1 and 2: '*P. aureus*' TSY, 3: '*Entothionella*'). (component 1: aurantioside A, 2: putative aurantioside derivative with shorter polyene length, 3: carotenoid, 4: protein, 5: cytochromes B and C, scale bar: 5 μm)

Figure 4: Insights into BGCs encoded in the ‘*P. aureus*’ genome. **a)** Organization of the auranoside (*ats*) gene cluster. **b)** Putative biosynthetic pathway for auranoside A (**2**). The assignments for AtsK and AtsN, marked with an asterisk (*), as sugar epimerases or

dehydratases are speculative. The residue modified by AtsJ is highlighted in orange. KS: ketosynthase, AT: acyl transferase, DH: dehydratase, KR: ketoreductase, C: condensation, A: adenylation, TE: thioesterase, H: halogenase, GT: glycosyl transferase, E: epimerase, NT: nucleotidyl transferase, DH: dehydratase, MT: methyl transferase, small black spheres: acyl/peptidyl carrier protein. **c)** Biochemical assays with the O-methyltransferase AtsJ. Extracted ion chromatograms are shown for authentic standards of **1** (red) and **2** (blue) (top) and for the conversion of **1** to **2** by AtsJ (bottom). SAM: S-adenosyl methionine. For control experiments, see figure S4. **d)** BGCs encoded in the 'P. aureus' hybrid draft genome. Triangles indicate stop-codons and frame shifts in the 'P. aureus' loci (*pau*) that could not be resolved by resequencing. KS: ketosynthase, CR: crotonyl-CoA reductase, A: adenylation, AT: acyl-transferase, DH: dehydratase, ER: enoyl reductase, KR: ketoreductase, MT: methyl transferase, Cy: cyclase, OX: oxidoreductase, C: condensation, TE: thioesterase, spheres: acyl/peptidyl carrier protein, GNAT: GCN5-related N-acetyltransferase, E: epimerase, ECH: enoyl-CoA hydratase, CAL: CoA ligase, Hy: hydrolase, DUF: domain of unknown function, ACD: acyl-CoA dehydrogenase, HCS: holocarboxylase synthase, Px: pyridoxal phosphate-dependent enzyme, AMT: amino transferase. Predicted substrates are noted below the respective A or AT domains.

Figure 5: Taxonomical and functional annotation of 'P. aureus' TSY. **a)** The phylogram was inferred based on concatenated marker genes by 1,000-bootstrap RAXML analysis. Concatenated sequences were generated from marker genes detected in all compared Chloroflexi genomes by CheckM. Color bar beside the branch labels shows the class of the genomes. **b)** Metabolic pathways of 'P. aureus' TSY as suggested by the presence of genes encoding transporters.

Figure. 6: SimGraph analysis of ABC transporter proteins of 'P. aureus' TSY. **a)** Orthologous transporter proteins (fasta36 E-value < 10^{-10}) were clustered. Colored dots show each transporter protein and its annotated function. **b)** The size distribution of ABC transporter clusters is shown against the cluster size rank in different bacteria.