Title

Synthase-Selective Exploration of a Tunicate Microbiome by Activity-Guided Single-Cell Genomics

Journal

ACS Chemical Biology, 16(5)

ISSN

1554-8929

Authors

Kim, Woojoo E

Charov, Katherine

Džunková, Mária

et al.

Publication Date

2021-05-21

DOI

10.1021/acschembio.1c00157

Peer reviewed

Letter

Synthase–selective exploration of a tunicate microbiome by activity–guided single–cell genomics

Woojoo E. Kim,[1,†] Katherine Charov,[1,†] Mária Džunková,[2,†] Eric D. Becraft,[3,4,†] Julia Brown,[3,†] Frederik Schulz,[2] Tanja Woyke,[2] James J. La Clair,[1] Ramunas Stepanauskas[3,*] & Michael D. Burkart[1,*]

[1] Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla CA 92093–0358, United States

[2] DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Mail Stop: 91R183, 1 Cyclotron Road, Berkeley, CA 94720, USA.

[3] Bigelow Laboratory for Ocean Sciences, East Boothbay ME 04544, United States

[4] University of North Alabama, Florence AL 35632, United States

[*] Corresponding authors: mburkart@ucsd.edu (M.D.B) and rstepanauskas@bigelow.org (R.S.)

[†] These authors contributed equally to the work.

Abstract. While thousands of environmental metagenomes have been mined for the presence of novel biosynthetic gene clusters, such computational predictions do not provide evidence of their *in vivo* biosynthetic functionality. Using a fluorescent *in situ* enzyme assay targeting carrier proteins common to polyketide (PKS) and non–ribosomal peptide synthetases (NRPS), we applied fluorescence–activated cell sorting to tunicate microbiome to enrich for microbes with active secondary metabolic capabilities. Single–cell genomics uncovered the genetic basis for a wide biosynthetic diversity in the enzyme–active cells and revealed a member of marine *Oceanospirillales* harboring a novel NRPS gene cluster with high similarity to phylogenetically distant marine and terrestrial bacteria. Interestingly, this synthase belongs to a larger class of siderophore biosynthetic gene clusters commonly associated with pestilence and disease. This demonstrates activity–guided single–cell genomics as a tool to guide novel biosynthetic discovery.

Introduction

Antecedent discoveries of secondary metabolites have been characteristically limited to their direct extraction from animals, plants, fungi and cultured microbes.[1] Nowadays, thousands of genomes of uncultured microbes from a variety of environmental samples are sequenced and mined for the presence of biosynthetic gene clusters (BGC), yet challenges still remain in the application of this approach to complex host–associated microbiomes containing yet uncultured bacteria.[2] Once a BGC is identified, the genes within these clusters can be synthesized and their biosynthetic enzymes expressed *in vitro* in culturable hosts (Figure S1 in the Supporting Information).[3] While an attractive approach to produce a secondary metabolite without culturing, considerable effort is required before one can reduce these large genomic datasets into actively

expressed biosynthetic gene clusters (BGCs) with associated secondary metabolite production.[4] One of the largest issues involves the discovery of a biosynthetic pathway appropriate for *in vitro* expression. While metagenome–based discoveries have revolutionized the way one can access microbes at the genetic level,[5] the presence of a biosynthetic gene cluster does not necessarily mean that it is functional *in vivo*.[6] Methods that enable sample selection based on biomarkers that confirm biosynthetic activity prior to genomic analyses could accelerate discoveries of bioactive molecules and increase the success with the associated bioactivity assays.

Here we apply fluorescent *in vivo* labeling of biosynthetic proteins as a tool to guide the selection of individual microorganisms expressing secondary metabolic activities of interest directly in their hosts for capture by FACS and single–cell genome sequencing. We demonstrate this approach on a microbiome from the tunicate *Ciona intestinalis* (Figure 1A). This activity–guided approach identifies targeted pathways in rare microorganisms without *a priori* knowledge of microorganism identity or cultivation.[7] As shown in Figure 1B, this workflow provides a robust complement to current genomic sequencing approaches by applying a fluorescent *in situ* biochemical readout as a tool for organism selection, a process that can be directly integrated into single–cell genomics workflows.[8]

To demonstrate this approach, we turned to carrier proteins (CPs) and their associated 4'–phosphopantetheinyltransferases (PPTases),[9] an enzyme–substrate pair that plays a key role in the biosynthesis of fatty acids (FAs), polyketides (PKs) and non–ribosomal peptides (NRPs). As shown in Figure 1B, this system provides a durable model, as we have previously shown that a variety of synthetic pantetheine analogues (pantetheinamides) can penetrate the cell for functional activity.[10] Once inside the cell, the pantetheinamide hijacks coenzyme A (CoA) biosynthesis for conversion to the corresponding CoA analog and ultimately becomes site–

selectively tethered onto the CP by the action of a PPTase (bottom, Figure 1B).[9] Importantly, the Sfp–type PPTases, which are associated with secondary metabolism of PKS and NRPS, are well known to show promiscuity for CoA identity, and therefore are excellent catalysts for labeling CPs *in vivo*. Here, we demonstrate how this process can be used to isolate bacterial cells engaged in active secondary metabolism (i.e., PKS and NRPS pathways) selectively over that of primary metabolism (fatty acid biosynthesis , FAS) by proper gating of a solvatochromic fluorescent pantetheinamide probe (DMN–P, Figure 1B).

Our study began by the selection of a dynamic fluorescent reporter. Developed in 2008,[11] 4–dimethylnaphthalene (4–DMN) is a solvatochromic tag with a fluorescence response dependent upon the hydrophobicity of its local environment.[12] In our previous studies, we used a 4–DMN–labeled pantetheinamide (DMN–P, Figure 1B) to probe protein–protein interactions between *Escherichia coli* fatty acid synthase (FAS) carrier protein (EcACP) and its multiple partner enzymes.[13] When an EcACP is labeled with DMN–P, the probe is sequestered within the hydrophobic alpha helical core of EcACP,[13] leading to a large increase in fluorescence intensity relative to that unbound in solution. Based on our previous studies,[10] we anticipated that DMN–P would cross the cell membranes and label CPs *in vivo*. The fact that the DMN–P was not fluorescent in solution but fluoresced when protein–bound suggested that this strategy could be used to directly identify cells containing the enzymatic machinery necessary to load a CP with DMN–P, as illustrated in Figure 1B. The use of CP–PPTase pairs was ideally suited to develop this method as CP domains are found in the majority PKS and NRPS systems.

Results and Discussion

We began the present study by establishing conditions to sort cells selectively based on the presence of an active synthase. While bacterial cells contained fatty acid synthases (FAS), we

postulated that FAS ACPs would be in a steady, active *holo*–state and would not label (note that the labeling method with results displayed in Fig. 1 only labels *apo*–CPs), while ACP/PCPs from NRPS/PKS biosynthesis would be expressed differentially over the bacterial cell cycle, therein allowing selective–labeling of their *apo*–states. Here, active NRPS/PKS cells could be sorted by setting a gate just above the levels of FAS ACP labeling. Using marine rod–shaped *Bacillus sp.* CNJ803 as a model, we identified conditions that provided a consistent cytosolic localization as evident in Figure 1C. This was then compared with a mutant deleted in Sfp, the 4'–phosphopantetheinyl transferase associated with the surfactin biosynthetic pathway in *B. subtilis*. Using identical conditions, we were able to establish selectivity (compare blue in Figure 1C to that of the deleted mutant in Figure 1D) and develop a gate for fluorescence–activated cell sorting (FACS) as shown in Figure 1E.

We then applied this protocol on a microbiome sampled from the colonial marine tunicate *Ciona intestinalis*, a model organism used in developmental biology, evolutionary biology, neuroscience,[14,15] and recently suggested as a model for microbially–associated secondary metabolism.[16] It was also one of the first animals to have its genome sequenced.[14] Adjacent tunicates specimens of *Ciona intestinalis* (Figure 1A) were collected from a dock in the Gulf of Maine. Freshly collected tunicates were incubated in sterile–filtered seawater along with control samples of the proximal water column (WC) with either RedoxSensor Green (RSG), a marker of bacterial cell viability, or the enzyme labeling DMN–P. To ensure sufficient sample of microbiome cells, a total of four tunicates were pooled for each probe. After homogenization of the tunicates, probe–positive cells were subjected to fluorescence–activated cell sorting (FACS, Figure 1E) to deposit fluorescently–labeled cells into a 384–well microplate, one per well (see Supporting Information).[17] Of the ~12,000 μL^{-1} viable (RSG–positive) microbial cells in its

tunicate homogenate, only ~13 μL^{-1} (0.1%) were labeled with DMN–P stained homogenate, suggesting high probe specificity. The DNA of 146–148 sorted particles per treatment was amplified, sequenced and *de novo* assembled individually (Table S1 in the Supporting Information).[17] In total, we obtained the following number of >20 kbp genome assemblies from individual, sorted particles: a) 50 of tunicate microbiome CP–expressing cells (DMN–P probe positives); b) 59 of tunicate microbiome respiring cells (RSG probe positives) representing the negative control without the probe; and c) 95 of adjacent seawater bacterioplankton (RGS probe positives).

The phylogenetic composition of the analyzed microbial cells from the adjacent water column (WC) was typical of the Gulf of Maine prokaryoplankton,[18] with a predominance of pelagic lineages SAR11 (*Alphaproteobacteria*) and *Flavobacteriales* (*Bacteroidetes*) (Figure 2A; Table S1 in the Supporting Information). The viable microorganisms in the tunicate microbiome were dominated by *Rhizobiales, Kordiimonadales* (*Alphaproteobacteria*), *Flavobacteriales* (*Bacteroidetes*) and *Campylobacteriales* (*Epsilonproteobacteria*), which have been previously shown to be associated with *C. intestinalis*.[19] The DMN–P probe enriched the tunicate microbiome for *Alteromonadales, Cellvibrionales, Vibrionales* and *Oceanospirillales* (*Gammaproteobacteria*), which were not detectable in the total active tunicate microbiome.

The specificity of the probe was further supported by the enrichment of sequences coding for PKS and NRPS synthases observed in the DMN–P sorted microbes (Figure 2B). As shown in Figure 2B, antiSMASH 5.0[20] identified 0.77 metabolite pathways per Mbp in the RSG probe–positive single amplified genomes (SAGs) and 1.17 metabolite pathways per Mbp in DMN–P positive SAGs, a level that was considerably higher than the 0.47 metabolite pathways per Mbp

that were observed in the adjacent water column (p–value 0.024). This observation confirmed that activity–based probes, such as DMN–P, can select microbes with biosynthetic potential.

Next, we explored two CPs from two different SAGs to provide evidence that the detected CP had been truly targeted by the DMN–P probe and to confirm the CP specificity. The first CP1, and its associated PPTase, PPT1, were obtained from a putative NRPS/Type I hybrid gene cluster (this gene cluster was observed only in this SAG) in an *Oceanospirillales* (*Gammaproteobacteria*) genome AH–491–C20 (red, Figure 2C), which had the highest 16S rRNA gene similarity (97%, 80% length overlap) to *Amphritea spongicola* MEBiC0546.[21] CP1 was chosen as it provided an excellent example of a hybrid NRPS–PKS, had a proximal PPTase, and it contained a unique domain architecture (Figure 2C) suggesting it was a novel NRPS. Analysis of tetramer frequencies and contig binning showed that the contig containing the NRPS genes was similar to other contigs with *Oceanospirillales* marker genes, which excludes the possibility that the NRPS gene cluster belonged to another co–sorted microbe or contaminating DNA (Figure S2 in the Supporting Information). A second, unrelated CP2, was obtained from a predicted NPRS/NRPS–like/Type I PKS cluster in a *Cellvibrionales* (*Gammaproteobacteria*) genome AH–491–D14, with *Oceanicoccus sagamiensis* NBRC107125 as the most closely related cultured isolate in Genbank (96% 16S rRNA gene identity, 81% overlap).[22] The NRPS–PKS CP2 was chosen, as it provided an excellent comparison with CP1.

Escherichia coli codon optimized genes were synthesized for CP1 (Figure S3 in the Supporting Information), CP2 (Figure S4 in the Supporting Information), and PPT1 (Figure S5 in the Supporting Information), inserted into pET28a vectors, and their associated proteins were prepared by recombinant expression in *E. coli* followed by His_6–tagged purification (Figure S6A in the Supporting Information). Applying the method in Figure 1B *in vitro*, samples of CP1 and

CP2 were screened for their ability to be fluorescently labeled. Recombinant Sfp,[9] a member of the surfactin–type PPTase known to have a broad scope in CP labeling, was able to load DMN–P onto CP1 under conditions established to label the EcACP, a positive control (Figure 2D, Figure S6A in the Supporting Information). Under the initial experimental conditions, CP2 was not labeled, nor was CP1 or CP2 labeled with the PPT1. Concerned that unfolding the proteins under SDS–PAGE would destroy the environmentally sensitive fluorescence from the DMN–P probe, we repeated the labeling process (Figure 1B) with a non–solvatochromic dye labeled CoA, TAMRA–CoA.[23] Here we observed labeling of CP1 with Sfp (Figure 2E, Figure S6B–D in the Supporting Information). Interestingly, PPT1 was only able to label CP1, which comes from the same bacterial species. Remarkably, while Sfp could label EcACP and CP1, the fact that PPT1 only labeled its native substrate CP1, and not CP2 or EcACP, demonstrates the unique selectivity found within PPTases.[9]

To explore the diversity and presence of biosynthetic gene clusters related to the *Oceanospirillales* bacterium AH–491–C20 NRPS across Bacteria and Archaea in publicly available genomic datasets, we used IMG/M database[24] (March 2020, containing 77,808 genomes from isolates, SAGs and metagenome assembled genomes from diverse environments) and 12,715 SAGs from the GORG–Tropics data set from marine prokaryoplankton.[25] Surprisingly, the AH–491–C20 NRPS gene cluster was not present in any known members of the *Oceanospirillales* order (Figure 2F), with the exception of *Zymobacter palmae,* an ethanol–fermenting species isolated from palm sap,[26] which had moderate amino acid sequence similarity (40–60%) to this gene cluster, but with a different domain organization (Figure 2C). An amino acid sequence similarity below 60% and different domain organization was also found in other members of *Proteobacteria* and *Firmicutes* phyla (Figure 2F). The most similar domain

organizations were found in pathogenic bacteria, such as *Vibrio cholerae* and *Vibrio mimicus*[27] and in marine bacteria, such as *Shewanella psychrophila* and *Photobacterium profundum*[28] (Figure 2C) belonging to different orders in the *Proteobacteria*. This yersiniabactin– and vibriobactin–like cluster was not found in any of the SAGs generated using the RSG probe in this study, which indicates the cluster's low abundance in the total active tunicate microbiome and highlights the utility of the activity–guided single–cell genomics for bioactive molecules.

Overall, this study demonstrates how activity–based fluorescent labeling can be coupled with single–cell genomics to enrich for organisms expressing specific biosynthetic activity. The usage of environmentally sensitive fluorescent probes designed to label CPs *in situ* enabled the identification of previously undiscovered CP substrate and compatible PPTase enzyme partners in the same, uncultured microbial cell. The use of this solvatochromic DMN–P probe provided an enhanced response (increased fluorescence when attached to a CP[9]) that enabled selective detection of bacteria presenting CP–containing synthases. Using model bacteria, we were able to show that the DMN–P probe selectively labeled microbial cells and we were able to use this signal to gate cell sorting for cells that would most likely contain an active NRPS or PKS CP–PPTase pair. While one cannot rule out that FAS pathways may have also been labeled, our data suggest that the application of sort gates with strict fluorescence intensity thresholds selects for microbial cells that contain added CP–PPTase activity associated with NRPS/PKS biosynthesis. We illustrated the enrichment for bacteria with biosynthetic activity with an example of a novel NRPS gene cluster found in an *Oceanospirillales* species from tunicate microbiome, which was below detection in the non–labeled control and also absent in large public repositories of microbial genomes. While well recognized in their human health context (see homology to *Yersinia pestis* and *Vibrio cholerae* in Figure 2C), the discovered yersiniabactin and vibriobactin

biosynthesis pathways have not been documented in *Oceanospirillales* nor tunicate microbiome previously. This indicates that such novel biochemistry would remain undiscovered without the use of our activity–guided single–cell genomics approach.

Methods

Probes and materials. Unless described otherwise, supplies and materials were obtained from VWR or Fischer Scientific and used as is. The DMN–P[13] and TAMRA–CoA[23] were prepared by chemical synthesis. *E. coli* FAS ACP (EcACP) was prepared using established methods.[13]

Microbial labeling studies. A single colony of *Bacillus sp.* CNJ803 (a marine wild type strain) or *Bacillus subtilis* 168 (a mutant deficient in Spf) was suspended in 200 μL of A1 media. Cells were then treated at 23 °C for 3 h with 500 nM DMN–P (Figure 1B) from a 100 μM stock solution of DMN–P dissolved in DMSO. Following incubation with DMN–P, nuclear stain, SYTO–9 (Thermo Fisher Scientific), was added to the cultures for 15 min at a final concentration of 1 μM. Cells were spun down at 2,000 rpm for 1 min and the supernatant was removed. Cell pellets were mixed with 200 μL of 3:1 EtOH/AcOH (fixing solution) and incubated for 10 min. Cells were centrifuged at 2,000 rpm for 1 min again and the supernatant was removed. The cells were re–hydrolyzed with 100 μL of water. A 20 μL aliquot of each final sample was loaded onto glass slides for super–resolution microscopy.

Super–resolution microscopy. Imaging was conducted on a Zeiss LSM 880 microscope with FAST Airyscan using a Plan–Apochromat 63x/1.4 Oil DIC M27 objective. Blue fluorescence from the DMN–P was obtained using a 405 nm laser with beamsplitters (MBS 488/561/633, MBS_InVis MBS -405, DBS1 mirror, FW1 rear) and an additional 410–477 nm filter. Green fluorescence from the SYTO–9 control was obtained using a 488 nm laser with beamsplitters (MBS 488/561/633, MBS_InVis MBS–405, DBS1 mirror, FW1 rear). Pinhole sizes were kept

between 40–90 μm. Images were collected in Zen (Zeiss) with recommended gains of 700–1100, digital gain of 1, depth of focus between 0.65–0.71 μm, and pixel time of 2–5 μs and image sizes at 2048 × 2048 pixels. Images were processed offline and rendered from Zen Blue (Zeiss). Copies of original CZI files can be provided upon request.

Specimen collection. Live *Ciona intestinalis* were collected from 20.4 °C seawater at a depth of ~0.5 m on the side of a floating dock in Boothbay Harbor Maine (Latitude 48.8, Longitude -69.6) between 8:00–9:00 AM on August 29, 2018. To ensure sufficient sample size, a total of four tunicates (these tunicates were collected from the same location at the same time and were neighboring and attached to the same substrate) were pooled for each experiment. Immediately after collection, the specimens were placed in 50 mL centrifuge tubes containing ambient seawater. Additionally, ambient seawater samples were collected adjacent to tunicates in 50 mL centrifuge tubes. The samples were transported to Bigelow Laboratory for Ocean Sciences at *in situ* temperature in the dark.

Activity–guided single–cell genomics. To target cells with CP activity, four tunicates were incubated in ambient seawater amended with 10 μM DMN–P at *in situ* temperature in the dark for 2 h. To target all viable microbial cells, four other tunicates were incubated in 80 mL of ambient seawater amended with 1 μM RedoxSensor Green (RSG, Thermo Fisher Scientific) at *in situ* temperature in the dark for 20 min. The tunicates of each treatment were rinsed in sterile Sargasso Sea water, pooled and homogenized together with 50 mL sterile Sargasso Sea water using a Ninja Professional 900 W blender until the majority of the biomass was visibly disintegrated. The homogenate was spun down at 2,000 rpm for 1 min and the supernatant was passed through a 100 μm mesh filter. To assess the composition of the sortable microorganisms in the seawater around tunicates, ambient seawater samples were labeled with a 5 μM SYTO–9

live nucleic acid stain (Thermo Fisher Scientific) for 20–40 min. Immediately before cell sorting, samples were diluted 10× in sterile Sargasso Sea water. Sort gates for DMN–P probe–positive cells were defined along blue fluorescence and forward scatter axes and adjusted for background noise using negative (*Bacillus subtilis* 168, a mutant deficient in Sfp) and positive (*Bacillus subtilis* 3610 DSM10, wild type) controls (culture conditions were the same as described in the above section entitled Microbial isolate labeling studies**)**. Probe–positive cells were sorted into 384–well plates, lysed, their DNA amplified with WGA–X, and genomes sequenced and quality–controlled.[17] Genome assemblies originating from multiple co–sorted cells were identified and parsed using a combination of nucleotide tetramer principal component analysis and homology searches in the NCBI nr database.[29] Biosynthetic pathways were identified with antiSMASH[20] using KnownClusterBlast, ActiveSiteFinder and SubClusterBlast options. If antiSMASH predicted multiple metabolite pathways in the same coding region, then all possible products were reported.

Heterologous expression of select genes/domains. Selected domains and di–domains were chosen based upon evaluation of sequencing data with antiSMASH (versions 4.2 and 5.0 were used).[20] In particular, we selected a PPTase (PPT1) and two CP (CP1 and CP2) containing sequences for recombinant evaluation. The predicted protein sequences of the full selected genes were further evaluated with BLAST alignment to identify homologues and better understand domain organization. In the case of CPs, disconnection locations for excised domains were determined by online domain organization tools[30,31] and by comparison with recent crystal structures.[32,33] The genes for the resulting proteins were synthesized (Twist Bioscience) and cloned into a pET28a vector. With the exception of the PPTase, all genes were cloned with a 3'–stop codon to provide an N–terminal His_6 tag for immobilized metal affinity purification. The

PPTase gene was cloned without a stop codon to provide a C–terminal His_6 tag based upon our prior experience with loss of activity with N–terminal fusions.[11] The identified protein sequences, gene sequences synthesized and protein sequences are provided in Figures S4–S6 in the Supporting Information.

Expression of CP1, CP2, and PPT1. The His_6–tagged proteins were expressed in *E. coli* (BL21), and grown in Terrific Broth. Cells were grown in the presence of 50 mg/L kanamycin, induced with 1 mM isopropyl β–D–1–thiogalactopyranoside (ITPG) at OD_{600} = 0.8, and incubated at 16 °C for 16 h. The cell culture was spun down by centrifugation at 2,000 rpm for 30 min and the collected pellets were lysed by sonication, followed by another centrifugation at 10,000 rpm for 1 h to clear the lysate. The proteins were purified using Ni–NTA resin (ThermoFisher). Purified proteins were collected and concentrated to 2–4 mg/mL using 3 kDa (CP1 and CP2) or 10 kDa (PPT1) Amicon Ultra centrifuge filters (Millipore).

Carrier protein labeling studies. The CP labeling studies were conducted in a 30 µL reaction containing final concentration of 100 µM of the respective CP1 or CP2, 1 mM of TAMRA–CoA, and 1 µM of Sfp or PPT1 in 50 mM HEPES, 10 mM $MgCl_2$ pH 7.2. The mixture was gently shaken at 23 °C overnight (12 h). The resulting reactions were analyzed by 12% SDS–PAGE gel and imaged on Typhoon TRIO Variable Mode Imager (GE Healthcare BioSciences). Imaging was conducted using Cy2 (Excitation 473 nm, Emission 520 nm) for DMN–P and Cy3 (Excitation 532 nm, Emission 580 nm) for TAMRA–CoA.

References

1. Kenshole, E., Herisse, M., Michael, M. and Pidot S. J. (2021) Natural product discovery through microbial genome mining. *Curr. Opin. Chem. Biol. 60*, 47–54.

2. Palaniappan, K., Chen, I.–M. A., Chu, K., Ratner, A., Seshadri, R., Kyrpides, N. C., Ivanova, N. N., and Mouncey, N. J. (2020) MG–ABC v.5.0: an update to the IMG/Atlas of biosynthetic gene clusters knowledgebase. *Nucleic Acids Res*. *48*, D422–D430.

3. Wang, G., Zhao, Z., Ke, J., Engel, Y., Shi, Y.–M., Robinson, D., Bingo, K., Zhang, Z., Bowen, B., Louie, K., Wang, B., Evans, R., Miyamoto, Y., Cheng, K., Kosina, S., De Raad, M., Silva, L., Luhrs, A., Lubbe, A., Hoyt, D. W., Francavilla, C., Otani, H., Deutsch, S., Washton, N. M., Rubin, E. W., Mouncey, N. J., Visel, A., Northen, T., Chen, J.–F., Bode, H. B., and Yoshikuni, Y. (2019) CRAGE enables rapid activation of biosynthetic gene clusters in undomesticated bacteria. *Nat. Microbiol. 4*, 2498–2510.

4. Kalkreuter, E., Pan, G., Cepeda, A. J., and Shen, B. (2020) Targeting bacterial genomes for natural product discovery. *Trends Pharmacol. Sci*. *41*, 13–26.

5. Miller, I. J., Chevrette, M. G., and Kwan, J. C. (2017) Interpreting microbial biosynthesis in the genomic age: biological and practical consideration. *Mar. Drugs 15*, 165.

6. Li, L., Liu, X., Jiang, W., and Lu, Y. (2019) Recent advances in synthetic biology approaches to optimize production of bioactive natural products in Actinobacteria. *Front. Microbiol*. *10*, 2467.

7. Martinez–Garcia, M., Brazel, D. M., Swan, B. K., Arnosti, C., Chain, P. S. G., Reitenga, K. G., Xie, G., Poulton, N. J., Gomez, M. L., Masland, D. E. D., Thompson, B., Bellows, W. K. B., Ziervogel, K., Lo, C.–C., Ahmed, S., Gleasner, C. D., Detter, C. J., and Stepanauskas, R. (2012) Capturing single cell genomes of active polysaccharide degraders: an unexpected contribution of *Verrucomicrobia*. *PLoS One 7*, e35314.

8. Doud, D. F. R., and Woyke, T. (2018) Novel approaches in function–driven single–cell genomics. *FEMS Microbiol. Rev*. *41*, 538–548.

9. Beld, J., Sonnenschein, E. C., Vickery, C. R., Noel, J. P., and Burkart, M. D. (2014) The phosphopantetheinyl transferases: catalysis of a post–translational modification crucial for life. *Nat. Prod. Rep. 31*, 61–108.

10. Reyes, C. P., La Clair, J. J., and Burkart, M. D. (2010) Metabolic probes for imaging endosymbiotic bacteria within toxic dinoflagellates. *Chem. Comm*. *46*, 8151–8153.

11. Loving, G., and Imperiali, B. (2008) A versatile amino acid analogue of the solvatochromic fluorophore 4–N,N–dimethylamino–1,8–naphthalimide: a powerful tool for the study of dynamic protein interactions. *J. Am. Chem. Soc. 130*, 13630–13638.

12. Klymchenko, A. S. (2017) Solvatochromic and fluorogenic dyes as environment–sensitive probes: Design and biological applications. *Acc. Chem. Res. 50*, 366–375.

13. Charov, K., and Burkart, M. D. (2019) A single tool to monitor multiple protein–protein interactions of the escherichia coli acyl carrier protein. *ACS Infect. Dis*. *5*, 1518–1523.

14. Satou, Y., Nakamura, R., Yu. D., Yoshida, R., Hamada, M., Fujie, M., Hisata K., Takeda, H., and Satoh, N. (2019) A nearly complete genome of *Ciona intestinalis* type a (C. *robusta*) reveals the contribution of inversion to chromosomal evolution in the genus Ciona. *Genome Biol. Evol*. *11*, 3144–3157.

15. Cao, C., Lemaire, L. A., Wang, W., Yoon, P. H., Choi, Y. A., Parsons, L. R., Matese, J. C., Wang, W., Levine, M., and Chen, K. (2019) Comprehensive single–cell transcriptome lineages of a proto–vertebrate. *Nature 571*, 349–354.

16. Leigh, B. A., Liberti, A., and Dishaw, L. J. (2016) Generation of germ–free *Ciona intestinalis* for studies of gut–microbe interactions. *Front. Microbiol*. *7*, 2092.

17. Stepanauskas, R., Fergusson, E. A., Brown, J., Poulton, N. J., Tupper, B., Labonté, J. M., Becraft, E. D., Brown, J. M., Pachiadaki, M. G., Povilaitis, T., Thompson, B. P.,

Mascena, C. J., Bellows, W. K., and Lubys, A. (2017) Improved genome recovery and integrated cell–size analyses of individual uncultured microbial cells and viral particles. *Nat. Commun*. *8*, 84.

18. Giovannoni, S. J. (2017) SAR11 Bacteria: The Most Abundant Plankton in the Oceans. *Ann. Rev. Mar. Sci. 9*, 231–255.

19. Blasiak, L. C., Zinder, S. H., Buckley, D. H. and Hill, R. T. (2014) Bacterial diversity associated with the tunic of the model chordate *Ciona intestinalis*. *ISME J. 8*, 309–320.

20. Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., Medema, M. H. and Weber, T. (2019) antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res*. *47,* W81–W87.

21. Jang, H., Wang, S.–H., Seo, H.–S., Lee, J.–H., Kim, S.–J., and Kwon K. K. (2015) *Amphritea spongicola* sp. nov., isolated from a marine sponge, and emended description of the genus Amphritea. *Int. J. Syst. Evol. Microbiol. 65*, 1866–1870.

22. Park, S., Yoshizawa, S., Kogure, K., and Yokota, A. (2011) *Oceanicoccus sagamiensis* gen. nov., sp. nov., a gammaproteobacterium isolated from sea water of Sagami Bay in Japan. *J. Microbiol*. *49*, 233–227.

23. Tallorin, L., Wang, J., Kim, W. E., Sahu, S., Kosa, N. M., Yang, P., Thompson, M., Gilson, M. K., Frazier, P. I., Burkart, M. D., and Gianneschi, N, C. (2018) Discovering de novo peptide substrates for enzymes using machine learning. *Nat. Commun*. *9*, 5253.

24. Chen, I. A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., Huntemann, M., Varghese, N., White, J. R., Seshadri, R., Smirnova, T., Kirton, E., Jungbluth, S. P., Woyke, T., Eloe–Fadrosh, E. A., Ivanova, N. N., and Kyrpides, N. K. (2019) IMG/M

v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res*. *47*, D666–D677.

25. Pachiadaki, M. G., Brown, J. M., Brown, J., Bezuidt, O., Berube, P. M., Biller, S. J., Poulton, N. J., Burkart, M. D., La Clair, J. J., Chisholm, S. W., and Stepanauskas, R. (2019) Charting the complexity of the marine microbiome through single–cell henomics *Cell*. *179*, P1623–1635.E11.

26. Okamoto, T., Taguchi, H., Nakamura, K., Ikenaga, H., Kuraishi, H., and Yamasato. K. (1993) *Zymobacter palmae* gen. nov., sp. nov., a new ethanol–fermenting peritrichous bacterium isolated from palm sap. *Arch. Microbiol. 160*, 333–337.

27. Keating, T. A., Marshall, C. G., and Walsh, C. T. (2000) Reconstitution and characterization of the Vibrio cholerae vibriobactin synthetase from VibB, VibE, VibF, and VibH. *Biochemistry. 39*, 15522–15530.

28. Oku, N., Kawabata, K., Adachi, K., Katsuta, A., and Shizuri, Y. (2008) Unnarmicins A and C, new antibacterial depsipeptides produced by marine bacterium *Photobacterium* sp. MBIC06485. *J. Antibiot. 61*, 11–17.

29. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol*. *215*, 403–410.

30. Khater, S., Gupta, M., Agrawal, P., Sain, N., Prava, J., Gupta, P., Grover, M., Kumar, N., and Mohanty, D. (2017) SBSPKSv2: structure–based sequence analysis of polyketide synthases and non–ribosomal peptide synthetases. *Nucleic Acids Res*. *45*, W72–W79.

31. Bachmann, B. O., and Ravel J. (2009) Chapter 8. Methods for *in silico* prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. *Methods Enzymol*. *458*, 181–217.

32. Miller, B. R., Sundlov, J. A., Drake, E. J., Makin, T. A., and Gulick, A. M. (2014) Analysis of the linker region joining the adenylation and carrier protein domains of the modular nonribosomal peptide synthetases. *Proteins 82*, 2691–2702.

33. Mitchell, C. A., Shi, C., Aldrich, C. C., and Gulick, A. M. (2012) Structure of PA1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains. *Biochemistry*. *51*, 3252–3563.

Acknowledgments

This study was supported by NIH R21 AI 134037 to M. D. B. and R. S. R. S. was also supported by NSF 826734, 1441717 and 133581. M. D. and T. W. were funded by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE–AC02–05CH11231.We thank W. Fenical (Scripps Institution of Oceanography, UC San Diego) for samples of *Bacillus sp.* CNJ803 used to develop the labeling conditions and J. Piel (ETH) for the *Bacillus subtilis* strains used as positive and negative controls. Confocal imaging was possible through the assistance of Jennifer Santini (UC San Diego) and supported by funding from the NIH P30 NS047101.

Author contributions

K. C. prepared the samples of DMN–P; W. E. K., K. C. and J. J. L. conducted the super–resolution imaging; E. D. B., J. B. and R. S. collected the tunicates and associated water column; E. D. B. and J. B. prepared the microbial extracts from the tunicates; E. D. B., J. B. and R. S. conducted the FACS analyses; E. D. B., J. B. and R. S. conducted the single–cell genomic sequencing; E. D. B. and R. S. conducted the genome assemblies and conducted biosynthetic

pathway searches; E. D. B., K. C., W. E. K., M. D. B., J. J. L. and R. S. mined and selected the CP and PPTase genes; K. C. and W. E. K. cloned and expressed the CP and PPTases; K. C. and W. E. K. performed the CP labeling studies; M. D. and F. S. performed sequence analysis of the NRPS gene cluster; and T. W., J. J. L, M. D. B. and R. S. organized the study. All authors contributed to the writing.

Supporting information

The Supporting Information is available free of charge at:

Figures S1–S7 and experimental methods for the DMN–P FACS analyses, single–cell DNA amplification and sequencing, SAG classification and assembly, and biosynthetic pathway analyses have been provided. The file also contains enlarged microscopic images, gene and protein sequences of CP1, CP2 and PPT1, and raw images of the SDS–PAGE gels. (PDF)

Table S1 (XLS)

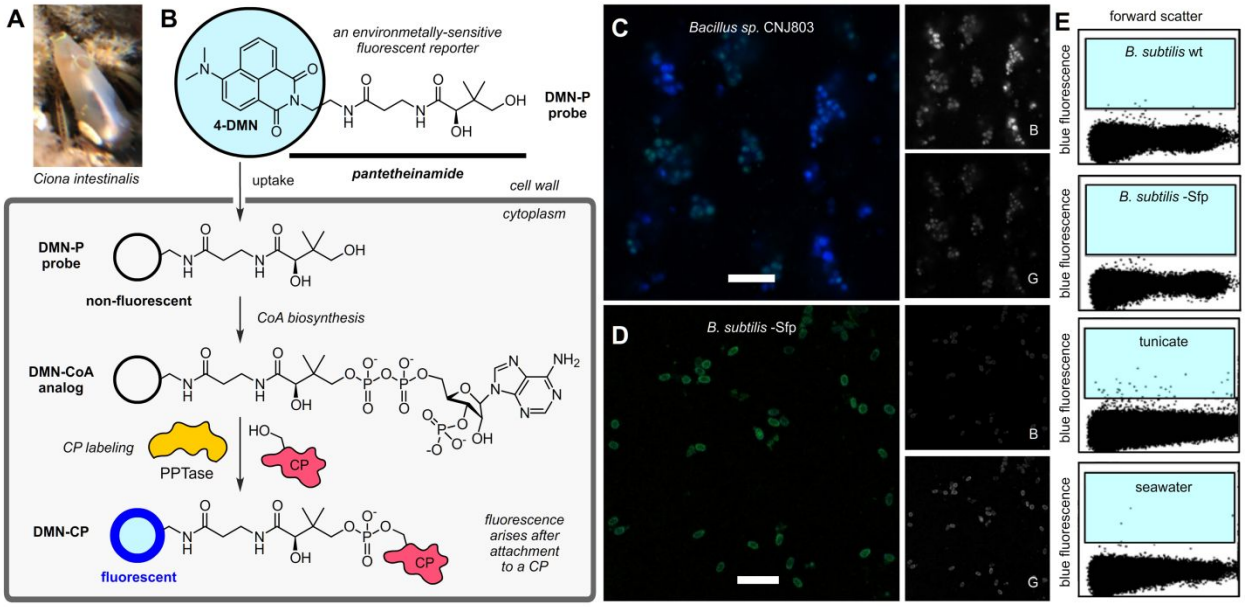

Figure 1 | Activity–guided microbial single–cell genomics. **A)** An image of the tunicate *Ciona intestinalis* specimen explored in this study. **B)** A schematic representation of the fluorescent CP labeling method used in this study. A fluorescent pantetheinamide DMN–P, designed to mimic pantetheine, is taken up by a microbial cell, where it is converted to a fluorescent–CoA analog. An environmentally sensitive fluorescent tag is used to improve the detection of protein–conjugation *in vivo*, as the 4–DMN dye used in the DMP–P probe has been shown to undergo an increase and shift in fluorescence once appended to a CP by a PPTase. **C)–D)** Super–resolution images of **C)** *Bacillus sp.* CNJ803 (a marine wild type strain) or **D)** *Bacillus subtilis* 168 (a mutant deficient in Sfp) incubated for 3 h with 500 nM DMN–P (blue) and then stained with 1 µM nuclear SYTO–9 (green) prior to fixation and imaging. Bar denotes 5 µm. **E)** Cells (black dots) sorted and selected (blue regions) by flow cytometry with DMN–P staining for tunicate and seawater samples. Negative (*Bacillus subtilis* 168, a mutant deficient in Sfp) and positive (*Bacillus subtilis* 3610 DSM10, wild type) controls indicate DMN–P stained cells with active CP • PPTase pairs. **(2 columns)**

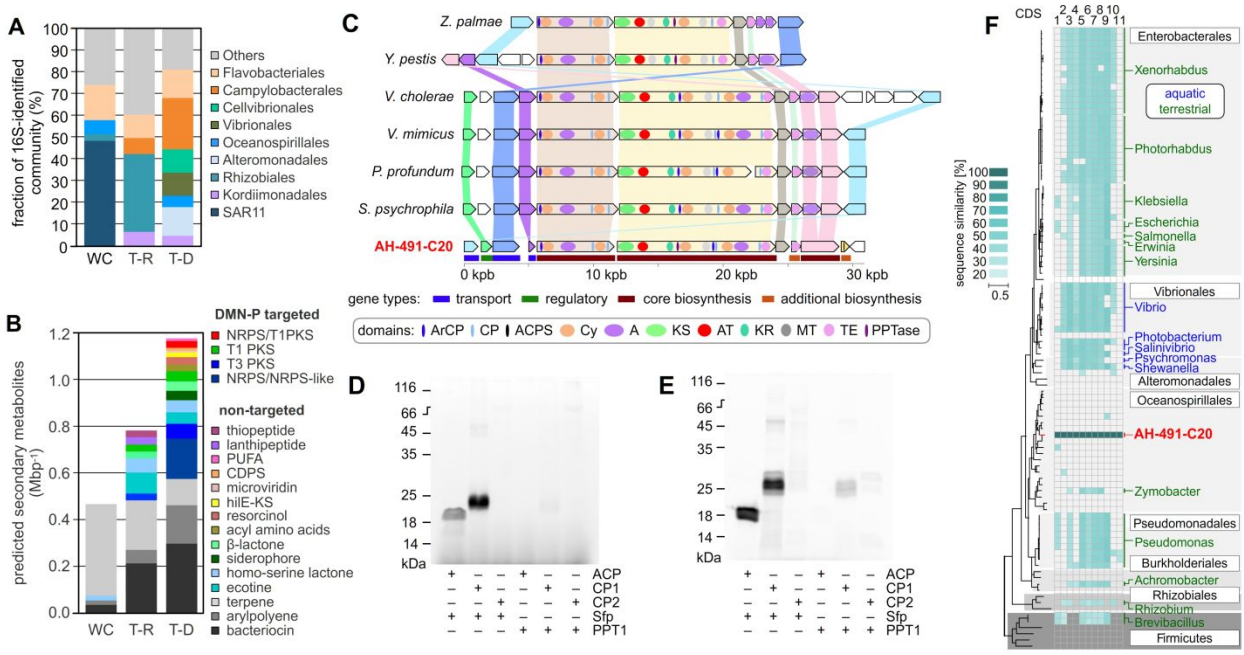

Figure 2 | Application of activity–guided single–cell genomics to mine selective CP and PPTase interactions from a tunicate microbiome. **A)** Bacterial composition and **B)** predicted secondary metabolite clusters per megabase (Mbp^{-1}) of the total microbial community in the water column (WC) compared to the viable (T–R) and DMN–P–responsive (T–D) cells in the *C. intestinalis* microbiome. *In vitro* CP–labeling analyses on the cloned and expressed CP1, CP2 and PPT1 mined from the genomic data. **C)** Visualization of the NRPS gene cluster from SAG AH–491–C20 and examples of alignments to other bacterial species which had >30% amino acid similarity on >75% of the query sequence length. **D)–E)** SDS–PAGE gel depicting the fluorescence in CPs after labeling by Sfp or PPT1 with **D)** DMN–P or **E)** TAMRA–CoA. *E. coli* FAS ACP (EcACP) was used as a positive control. **F)** Presence of the AH–491–C20 NRPS gene cluster across different phylogenetic groups. The phylogenetic tree includes genomes, which contained at least 7 out of 11 genes of this cluster and their closest relatives. Heatmap illustrates the level of sequence similarity for each gene. **(2 columns)**