Localized production of defence chemicals by intracellular symbionts of Haliclona sponges

Ma. Diarey Tianero, Jared N. Balaich and Mohamed S. Donia

Marine sponges often house small-molecule-producing symbionts extracellularly in their mesohyl, providing the host with a means of chemical defence against predation and microbial infection. Here, we report an intriguing case of chemically mediated symbiosis between the renieramycin-containing sponge Haliclona sp. and its herein discovered renieramycin-producing symbiont Candidatus Endohaliclona renieramycinifaciens. Remarkably, Ca. E. renieramycinifaciens has undergone extreme genome reduction where it has lost almost all necessary elements for free living while maintaining a complex, multi-copy plasmid-encoded biosynthetic gene cluster for renieramycin biosynthesis. In return, the sponge houses Ca. E. renieramycinifaciens in previously uncharacterized cellular reservoirs (chemobacteriocytes), where it can acquire nutrients from the host and avoid bacterial competition. This relationship is highly specific to a single clade of Haliclona sponges. Our study reveals intracellular symbionts as an understudied source for defence chemicals in the oldest-living metazoans and paves the way towards discovering similar systems in other marine sponges.

Marine sponges are among the richest metazoan hosts of microbial diversity. They are considered to be ancestral organisms and are a good model system for studying the evolution of microorganism–host interactions. Having evolved mechanisms to be recognized as symbiotic partners instead of food, or to avoid host detection altogether, most sponge symbionts are housed in their mesohyl matrix. These symbionts include those that provide nutrition for the host and help with waste recycling, provide structural elements such as calcium or phosphorus, and produce potent molecules as defence chemicals against predators and infectious agents. The latter is best exemplified by the filamentous bacterial symbiont of Theonella sponges, Candidatus Entotheonella sp., which produces numerous complex metabolites with different biological activities, and the cyanobacterial symbiont of Dysidea sponges, Hormoscilla spongiae, which produces several halogenated toxins. These two remarkable examples and the fact that more than 8,000 small molecules have been isolated from marine sponges suggest that microbiome-derived biosynthesis of sponge metabolites is a widespread phenomenon that is largely unexplored.

Examples of such understudied sponge metabolites include the renieramycins: a group of tetrahydroisoquinoline quinones (THQs) originally isolated from marine sponges of the genus Haliclona (previously known as Reniera), and subsequently from Xestospongia and Neopetrosia sponges. Approximately 30 renieramycins have been discovered to date, exerting a wide range of antimicrobial and cytotoxic activities that are relevant both ecologically and therapeutically. These biological activities suggest that the renieramycins may have defensive or competitive allelopathic roles that benefit the sponge host, although further studies are needed to fully understand their ecological role. From a biosynthetic point of view, much less is known about the renieramycins. The renieramycin core structure resembles that of the saframycin group of THQs, members of which are produced by disparate bacterial origins: free-living terrestrial myxobacteria and Actinobacteria (the saframycins), free-living Proteobacteria (the safracins) and obligate symbionts of marine ascidians (the euteinascidins), which include the recently approved anticancer drug ET-743 (also known as trabectedin). Intrigued by the widespread nature of this group of molecules, we set out to identify the source and molecular bases of renieramycin production in the most primitive metazoans: marine sponges.

**Results**

**Discovery of the renieramycin biosynthetic gene cluster.** Because of the structural similarity between the renieramycins and other THQs, we hypothesized that they are biosynthesized by a bacterial symbiont of Haliclona sponges through a non-ribosomal peptide synthetase (NRPS) pathway. To test this hypothesis, we studied four renieramycin-containing Haliclona sponge samples from different years of collection and geographical locations in the Tropical Pacific Ocean: Ren-Pal-02 (Palau, 2002), Ren-PNG-07060, Ren-PNG-07113 (Papua New Guinea, 2007) and Ren-Bali-16-03 (Bali, 2016) (Fig. 1). First, we used chemical analysis, high-performance liquid chromatography coupled with high-resolution tandem mass spectrometry (HPLC–HR-MS/MS) and nuclear magnetic resonance (NMR) to verify the presence of renieramycins in these sponges. In all samples, renieramycin E was the most prominent product renierone and its derivative, accompanied by the typical renieramycin degradation resonance (NMR) to verify the presence of renieramycins in these sponges. In all samples, renieramycin E was the most prominent derivative, accompanied by the typical renieramycin degradation products renierone and N-formyl dehydrorenierone (Fig. 1c and Supplementary Figs. 1 and 2). Second, we molecularly identified the four sponges (collected as Reniera or Haliclona based on their morphology) by amplifying and sequencing ~500 bp of their 18S rRNA gene. As expected, the four sequences were 99.9% identical to each other and to previously characterized Haliclona sponges.

Having confirmed the presence of renieramycins in our Haliclona sponges, we undertook an unbiased strategy towards identifying the renieramycin biosynthetic gene cluster (BGC) in their microbiomes. We isolated metagenomic DNA from the four sponges and subjected them to deep metagenomic sequencing using Illumina (30–110 million paired-end reads each of 141 bp or 175 bp in length; see Methods and Supplementary Table 1). Several rounds of assembly yielded major scaffolds with lengths up to ~700 kb. To identify the renieramycin BGC, we took two approaches: a targeted approach, in which we used tBLASTn to search the final assemblies (scaffolds > 5 kb)
of the four metagenomes for homologues of the saframycin Mx1 NRPSs SafA and SafB, and an untargeted approach, in which we analysed them using antiSMASH (a specialized tool for automatic identification of small-molecule BGCs). Remarkably, this analysis revealed that the four assembled Halichona metagenomes have very limited biosynthetic capacity, unlike previously studied sponge metagenomes. Overall, antiSMASH detected a single BGC from each of Ren-Pal-02, Ren-PNG-07060 and Ren-PNG-07113 and three BGCs from Ren-Bali-16-03 (Supplementary Fig. 3). Only one BGC was common to the four metagenomes: NRPS-1, which was found on a ~33-kb scaffold with >99% DNA sequence identity between the samples. Interestingly, the same BGC was also retrieved using the targeted saframycin Mx1-based search strategy and encodes a complete saframycin-like BGC: ren (Fig. 2a). Analogous to previously characterized THQ BGCs, ren encodes all of the enzymes necessary for the formation of the pentacyclic core of the molecule as well as substrate-modifying and tailoring enzymes, allowing us to propose a biosynthetic scheme for the renieramycins (Fig. 2, Supplementary Table 2 and Supplementary Note). To unequivocally demonstrate that ren encodes the renieramycins, we cloned and sought to heterologously express the entire ~25-kb ren scaffold to its matching scaffolds in the four samples had an average pairwise identity between the samples. Interestingly, the same BGC was also retrieved using the targeted saframycin Mx1-based search strategy and encodes a complete saframycin-like BGC: ren (Fig. 2a). Analogous to previously characterized THQ BGCs, ren encodes all of the enzymes necessary for the formation of the pentacyclic core of the molecule as well as substrate-modifying and tailoring enzymes, allowing us to propose a biosynthetic scheme for the renieramycins (Fig. 2, Supplementary Table 2 and Supplementary Note). To unequivocally demonstrate that ren encodes the renieramycins, we cloned and sought to heterologously express the entire ~25-kb ren scaffolds despite its high coverage (232–335× coverage). We then wondered whether ren exists on an extrachromosomal element. Indeed, analysis of paired-end reads spanning the ~33-kb scaffolds proved that they are joined end to end, and that ren exists on a small bacterial plasmid in all four metagenomes: p-ren (Fig. 3). Although very interesting from an evolutionary standpoint, this fact presented us with a technical challenge to unequivocally associate ren to specific members of the Halichona microbiome.

To identify the bacterium that harbours p-ren, we undertook two independent approaches: a computational approach and an experimental approach. First, based on the extreme sequence conservation between p-ren in the four metagenomes, we hypothesized that the genomes of the bacterium harbouring it would also be conserved and present at a similar coverage. To test this hypothesis, we identified all scaffolds that are in common between the four metagenomes by mapping the reads from Ren-PNG-07060, Ren-Pal-02 and Ren-PNG-07113 to the assembled scaffolds of Ren-Bali-16-03. Only 43 of 6,013 scaffolds (0.7%) are shared between the Ren-Bali-16-03 metagenome and the other three metagenomes. Satisfyingly, 40 of these scaffolds originated from the sponge genome, one corresponded to the sponge mitochondrion (a circularized 18,605 bp), one corresponded to p-ren and only the final one appeared to be of an unknown prokaryotic origin (Ren-Bali-16-03-NODE_1, ~680 kb). Ren-Bali-16-03-NODE_1 had three notable features: (1) its matching scaffolds in the four samples had an average pairwise nucleotide identity of 99.73–99.98%, (2) the coverage ratio of this scaffold to p-ren is 1/4.4, 1/3, 1/1.7 and 1/2.6 for Ren-PNG-07060, Ren-PNG-07113, Ren-Pal-02 and Ren-Bali-16-03, respectively (calculated by computing their RPKM (number of mapped reads per kilobase per million of sequenced reads) values), and (3) its GC content is similar to that of p-ren (38% versus 34%, respectively). Remarkably, paired-end read analysis circularized Ren-Bali-16-03-NODE_1 and its corresponding scaffolds into complete bacterial chromosomes of an extremely small size (~680 kb) (Fig. 3). Taken together, these results show that only one bacterial chromosome is consistently found in p-ren-containing samples at a sequence conservation level, coverage and GC content similar to that of p-ren, motivating us to hypothesize that these two genetic elements are harboured by the same bacterial cell. We propose to name this bacterium Ca. Endohalicona renieramycinicinaceae.

To test this hypothesis experimentally, we wondered whether the two elements can be physically colocalized. Inspired by elegant studies in Theonella sponges, we gently homogenized fresh tissues of Ren-Bali-16-03 and subjected them to flow cytometry and cell sorting, guided only by particle size (Supplementary Fig. 5). Overall, we sorted the homogenate into eight partitions (P3–P10), isolated DNA from each and performed high-throughput 16S rRNA gene amplification sequencing on all of them and metagenomic sequencing on a selected subset that had widely varying levels of Ca. E. renieramycinicinaceae (P4 and P7–P10) (Fig. 4a and Supplementary Table 1). Surprisingly, the relative abundance of the 16S rRNA gene sequence of Ca. E. renieramycinicinaceae was most enriched in the last partition (P10), which contained the largest particle size of the sponge...
Fig. 2 | Renieramycin biosynthesis. a, Renieramycin BGCs (ren) discovered from the four Haliclona sponges in this study (top) and previously characterized THQ BGCs from other microorganisms (bottom). b, Comparison of the NRPS domain architecture between ren and related BGCs. A, adenylation; ACL, acyl-coenzyme A ligase; C, condensation; T, thiolation; TD, terminal reductase. c, Chemical structures of the products encoded by the BGCs in b, showing the common pentacyclic core of the molecules in blue. Note that renieramycin E is one amino acid shorter than the typical molecules in this class, safracins and saframycins (nitrogen atoms of individual amino acids are shown in red), which agrees with ren missing the first A domain. d, Proposed biosynthesis of renieramycin E based on characterized homologues from this study and previous ones. PCP, peptidyl carrier protein. e, Extracted ion chromatograms (HPLC–HR-MS) monitored at m/z = 196.0967 for the following samples (from top to bottom): an authentic standard of O-methyl-L-tyrosine, an authentic standard of 3-methyl-L-tyrosine, an organic extract generated from the supernatant of E. coli cells expressing renB and supplemented with L-tyrosine, and an organic extract generated from the supernatant of E. coli cells harbouring an empty vector and supplemented with L-tyrosine. This experiment was repeated independently three times, each in a triplicated setup, and produced the same results.
homogenate. This enrichment is corroborated in the metagenomic sequencing data and corresponds to an increase in the coverage of both the Ca. E. renieramycinifaciens chromosome and p-ren (Fig. 4b,c and Supplementary Table 3). As expected, a covariance analysis between the coverage of the Ca. E. renieramycinifaciens chromosome and p-ren in the five partition metagenomes revealed a strong and positive correlation: Pearson correlation coefficient of 0.99, \( P = 4.08 \times 10^{-7} \). This covariance is maintained at an average RPKM ratio of 1/2.4 (Ca. E. renieramycinifaciens chromosome/p-ren), which agrees with their ratio in the full Ren-Bali-16-03 metagenome (1/2.6). In addition, the overall compositional complexity of the partition metagenomes decreases as the particle size increases, as shown by GC content versus coverage binning analyses (Supplementary Fig. 6). These results not only provide strong evidence that the plasmid and chromosome are colocalized but also that they are colocalized with sponge cells in the largest particles of the entire homogenate.

**Intracellular localization of Ca. E. renieramycinifaciens.** We reasoned that there are two main possibilities for the colocalization of Ca. E. renieramycinifaciens and p-ren in large particles: (1) Ca. E. renieramycinifaciens cells are clumped in a colony-like shape that is not easily disrupted, or (2) Ca. E. renieramycinifaciens cells exist intracellularly in an enlarged sponge cell (bacteriocyte). To differentiate between these two possibilities, we performed fluorescence in situ hybridization (FISH) on Ren-Bali-16-03 using both a mixture of general eubacterial probes (EU338 I, II and III) and a Ca. E. renieramycinifaciens-specific probe designed here (CE75), in addition to \( 4',6\text{-diamidino-2-phenylindole} \) (DAPI) staining for DNA visualization. The sponge tissue contains a heterogeneous population of cells, including typical sponge cells of 2–5 \( \mu m \) in diameter and noticeably large spherical, bacteriocyte-like cells of 15–25 \( \mu m \) in diameter, as well as siliceous sponge spicules (Supplementary Fig. 7).

Interestingly, the bacteriocyte-like cells, but not other sponge cells, appear packed with smaller DAPI-stained particles. Hybridization using universal eubacterial probes showed strong fluorescent signals from the bacteriocyte-like cells only, indicating that the smaller particles that fill them are indeed bacterial cells. Hybridization with a Ca. E. renieramycinifaciens-specific probe also showed localized signals in the bacteriocyte-like cells. Moreover, a composite image and colocalization microscopy analyses revealed that signals from the eubacterial and Ca. E. renieramycinifaciens-specific probes

---

**Fig. 3 | Ca. E. renieramycinifaciens genomes and plasmids.** Assembled circular chromosomes of Ca. E. renieramycinifaciens from the four sponge metagenomes and the corresponding renieramycin gene cluster containing plasmids, p-ren. Concentric rings (from outside to inside) indicate genes on the forward frame, genes on the reverse frame, RNAs, GC content and GC skew. Genes are classified according to general COG categories in the IMG ([A], RNA processing and modification; [B], chromatin structure and dynamics; [C], energy production and conversion; [D], cell cycle control, cell division and chromosome partitioning; [E], amino acid transport and metabolism; [F], nucleotide transport and metabolism; [G], carbohydrate transport and metabolism; [H], coenzyme transport and metabolism; [I], lipid transport and metabolism; [J], translation, ribosomal structure and biogenesis; [K], transcription; [L], replication, recombination and repair; [M], cell wall/membrane/envelope biogenesis; [N], cell motility; [O], post-translational modification, protein turnover and chaperones; [P], inorganic ion transport and metabolism; [Q], secondary metabolites biosynthesis, transport and catabolism; [R], general function prediction only; [S], function unknown; [T], signal transduction mechanisms; [U], intracellular trafficking, secretion and vesicular transport; [V], defence mechanisms; [W], extracellular structures; [X], mobilome, prophages and transposons; [Y], nuclear structure; [Z], cytoskeleton; [NA], not assigned). The colour code of genes in p-ren follows the same key in Fig. 2a.
mostly superimpose in the bacteriocyte-like cells (Pearson correlation coefficient of 0.94) (Fig. 5a). Although signals corresponding to the universal eubacterial probes can be observed elsewhere in the sponge tissue (other bacteria), Ca. E. renieramycinifaciens-specific signals are exclusively detected in the bacteriocyte-like cells (Supplementary Fig. 7). Finally, transmission electron microscopy (TEM) images of the bacteriocyte-like cells confirmed that they harbour what appears to be a single morphotype of bacterial cells (Fig. 5b). Altogether, these hybridization and microscopy experiments demonstrate that Ca. E. renieramycinifaciens is an intracellular symbiont that resides solely in large sponge bacteriocytes.

Although the flow cytometry, cell sorting and metagenomic binning analyses provided strong support for the colocalization of p-ren with the Ca. E. renieramycinifaciens chromosome, and in light of the microscopic revelation that Ca. E. renieramycinifaciens resides in specialized sponge bacteriocytes, we sought to perform one final experiment that definitively localizes p-ren to Ca. E. renieramycinifaciens. Aided by the easily recognizable spherical shape of the large sponge bacteriocytes, we used laser capture microdissection (LCM) to isolate 100 single cells of their kind from diluted Ren-Bali-16-03 sponge homogenates. We then isolated and sequenced DNA from the captured cells and analysed their metagenomes in comparison to DNA isolated from captured background membranes as a control. Finally, we quantified the number of reads that mapped to Ca. E. renieramycinifaciens genomes are synonymous with the same abundance profiles of genes in all pathways, indicating that they are mostly identical. Consistent with its intracellular lifestyle, the Ca. E. renieramycinifaciens genome shows several hallmarks of reduction: (1) it encodes all genes for DNA replication, but no genes for DNA repair, recombination or transposable and mobile elements; (2) central metabolic pathways (that is, glycolysis, pentose phosphate and tricarboxylic acid cycle) are incomplete; (3) except for a partial biosynthetic pathway for lysine, it lacks all pathways for amino acid, cofactor, prosthetic group, nucleoside and nucleotide biosynthesis; and (4) genes for cell surface structures, chemotaxis, toxin production and detoxification, and signal transduction are mostly absent. The loss of all of these functions further establishes Ca. E. renieramycinifaciens as an obligate long-term symbiont incapable of free living (Supplementary Table 4).

**Extreme reduction of the Ca. E. renieramycinifaciens genome.** Based on a phylogenetic tree constructed from the 16S rRNA gene, Ca. E. renieramycinifaciens sequences derived from this study and three closely related sequences (~99% identical to Ca. E. renieramycinifaciens (Fig. 5b)) obtained previously from Australian Haliclona sp. form a distinct clade within the class Gammaproteobacteria, with members from the families Coxiellaceae and Legionellaceae (order Legionellales) as the nearest neighbours (Supplementary Fig. 8). To gain more insights into their biochemical and metabolic capabilities, we annotated the Ca. E. renieramycinifaciens genomes using the Integrated Microbial Genome platform (IMG; http://img.jgi.doe.gov) and compared their encoded functions to that of Coxiella HT99, the closest free-living relative with a sequenced genome (89% 16S rDNA sequence identity) (Supplementary Table 4).

The four Ca. E. renieramycinifaciens genomes are syntenic and share an average nucleotide sequence identity of 99.98%. Moreover, an overview of all metabolic functions using TIGRFAM pathways shows that the four genomes share the same abundance profiles of genes in all pathways, indicating that they are mostly identical. Consistent with its intracellular lifestyle, the Ca. E. renieramycinifaciens genome shows several hallmarks of reduction: (1) it encodes all genes for DNA replication, but no genes for DNA repair, recombination or transposable and mobile elements; (2) central metabolic pathways (that is, glycolysis, pentose phosphate and tricarboxylic acid cycle) are incomplete; (3) except for a partial biosynthetic pathway for lysine, it lacks all pathways for amino acid, cofactor, prosthetic group, nucleoside and nucleotide biosynthesis; and (4) genes for cell surface structures, chemotaxis, toxin production and detoxification, and signal transduction are mostly absent. The loss of all of these functions further establishes Ca. E. renieramycinifaciens as an obligate long-term symbiont incapable of free living (Supplementary Table 4).

**Fig. 4 | p-ren and Ca. E. renieramycinifaciens colocalize with the largest sponge particles.** a. A schematic representation of the flow cytometry experiment and subsequent analyses performed on Ren-Bali-16-03. b. Relative abundance of the Ca. E. renieramycinifaciens 16S rRNA gene sequence in all eight flow partitions. Note that the relative abundance of Ca. renieramycinifaciens increases in later partitions containing larger particles (P9 and P10). c. Coverage (measured in RPKM) of the Ca. E. renieramycinifaciens chromosome and p-ren in the five partitions analysed by metagenomic sequencing. Note that the coverage of both genetic elements also increases in later partitions containing larger particles (P9 and P10).

Extreme reduction of the Ca. E. renieramycinifaciens genome. Based on a phylogenetic tree constructed from the 16S rRNA gene, Ca. E. renieramycinifaciens sequences derived from this study and three closely related sequences (~99% identical to Ca. E. renieramycinifaciens (Fig. 5b)) obtained previously from Australian Haliclona sp. form a distinct clade within the class Gammaproteobacteria, with members from the families Coxiellaceae and Legionellaceae (order Legionellales) as the nearest neighbours (Supplementary Fig. 8). To gain more insights into their biochemical and metabolic capabilities, we annotated the Ca. E. renieramycinifaciens genomes using the Integrated Microbial Genome platform (IMG; http://img.jgi.doe.gov) and compared their encoded functions to that of Coxiella HT99, the closest free-living relative with a sequenced genome (89% 16S rDNA sequence identity) (Supplementary Table 4).

The four Ca. E. renieramycinifaciens genomes are syntenic and share an average nucleotide sequence identity of 99.98%. Moreover, an overview of all metabolic functions using TIGRFAM pathways shows that the four genomes share the same abundance profiles of genes in all pathways, indicating that they are mostly identical. Consistent with its intracellular lifestyle, the Ca. E. renieramycinifaciens genome shows several hallmarks of reduction: (1) it encodes all genes for DNA replication, but no genes for DNA repair, recombination or transposable and mobile elements; (2) central metabolic pathways (that is, glycolysis, pentose phosphate and tricarboxylic acid cycle) are incomplete; (3) except for a partial biosynthetic pathway for lysine, it lacks all pathways for amino acid, cofactor, prosthetic group, nucleoside and nucleotide biosynthesis; and (4) genes for cell surface structures, chemotaxis, toxin production and detoxification, and signal transduction are mostly absent. The loss of all of these functions further establishes Ca. E. renieramycinifaciens as an obligate long-term symbiont incapable of free living (Supplementary Table 4).

**Fig. 4 | p-ren and Ca. E. renieramycinifaciens colocalize with the largest sponge particles.** a. A schematic representation of the flow cytometry experiment and subsequent analyses performed on Ren-Bali-16-03. b. Relative abundance of the Ca. E. renieramycinifaciens 16S rRNA gene sequence in all eight flow partitions. Note that the relative abundance of Ca. renieramycinifaciens increases in later partitions containing larger particles (P9 and P10). c. Coverage (measured in RPKM) of the Ca. E. renieramycinifaciens chromosome and p-ren in the five partitions analysed by metagenomic sequencing. Note that the coverage of both genetic elements also increases in later partitions containing larger particles (P9 and P10).
Reniermacyin biosynthesis requires several substrates and cofactors; remarkably, almost none of which seems to be produced by *Ca. E. renieramycinifaciens*. These include direct substrates, namely, angelic acid and tyrosine, and cofactors needed for the enzymatic activity, such as (R)-4′-phosphopantothenate, S-adenosyl methionine and nicotinamide adenine dinucleotide. Because of the
After data processing using the Qiime pipeline, we searched all assignments, whereas 7 are unclassified (Supplementary Table 5). The 1,194 GSM sponge metagenomes belonged to 96 types of sponges, 89 of which have genus-level or species-level names. Of the 1,435 marine samples belonging to 1,194 sponges, 37 clades of sponges were represented by multiple individuals, their average composition is shown (see Methods and Supplementary Table 5).

**Discussion**

The *Halichondra–Ca. E. renieramycinifaciens* system described here differs from other examples of defensive symbiosis in marine sponges in several aspects. First, unlike *H. spongidae* and *H. vanosui* symbioses, *Ca. E. renieramycinifaciens* is an intracellular symbiont of a specific clade of marine sponges, where it dominates the bacterial microbiome. The bacterial composition of the rest of the microbiome is shown at the phylum level, following the colour code on the left.

**A specific symbiosis between Ca. E. renieramycinifaciens and a clade of Halichondra sponges.** To gain insights into the distribution of *Ca. E. renieramycinifaciens* in the ocean, we compared the bacterial community composition of our four *Halichondra* sponges with that of 1,435 marine samples belonging to 1,194 sponges, 37 sediment metagenomes, 195 seawater metagenomes and 9 unclassified metagenomes from a Global Sponge Microbiome (GSM) analysis using high-throughput 16S rRNA gene sequencing (see Methods). The 1,194 GSM sponge metagenomes belonged to 96 types of sponges, 89 of which have genus-level or species-level assignments, whereas 7 are unclassified (Supplementary Table 5). After data processing using the Qiime pipeline, we searched all samples for operational taxonomic units (OTUs) matching *Ca. E. renieramycinifaciens* and mapped their distribution to a phylogenetic tree of the host sponges that we constructed based on their 18S rRNA gene sequences. Overall, we discovered OTUs matching *Ca. E. renieramycinifaciens* exclusively in sponge samples and not in the seawater or sediment samples included, supporting its obligate symbiotic lifestyle. Importantly, sequences matching *Ca. E. renieramycinifaciens* exist specifically in one clade of *Halichondra* sponges, namely, the four from this study and seven ‘blue Halichondras’ from the Pacific Ocean (Fig. 6), and is absent in all other sponges including *Halichondra* individuals that belong to other clades. Finally, when present, the relative abundance of *Ca. E. renieramycinifaciens* in *Halichondra* sponges is unusually high, ranging from 10% to 75% of the total 16S rRNA gene sequences in a given sample (Fig. 6 and Supplementary Fig. 9). These results establish *Ca. E. renieramycinifaciens* as an intracellular symbiont of a specific clade of marine sponges, where it dominates the bacterial community and specializes in producing the renieramcyins.

**Fig. 6 | Host specificity of *Ca. E. renieramycinifaciens*.** 18S rRNA gene-based phylogenetic tree of representative sponges (left) and the 16S rRNA gene-based bacterial composition of their corresponding microbiomes (right) (samples are from this study and the GSM). The phylogenetic tree was constructed using Fasttree-ML (Jukes–Cantor model), and local support values as fractions of 1,000 resamples are shown. When sponge species from the GSM are compared to the 18S rRNA gene sequences. Overall, we discovered OTUs matching *Ca. E. renieramycinifaciens* exclusively in sponge samples and not in the seawater or sediment samples included, supporting its obligate symbiotic lifestyle. Importantly, sequences matching *Ca. E. renieramycinifaciens* exist specifically in one clade of *Halichondra* sponges, namely, the four from this study and seven ‘blue Halichondras’ from the Pacific Ocean (Fig. 6), and is absent in all other sponges including *Halichondra* individuals that belong to other clades. Finally, when present, the relative abundance of *Ca. E. renieramycinifaciens* in *Halichondra* sponges is unusually high, ranging from 10% to 75% of the total 16S rRNA gene sequences in a given sample (Fig. 6 and Supplementary Fig. 9). These results establish *Ca. E. renieramycinifaciens* as an intracellular symbiont of a specific clade of marine sponges, where it dominates the bacterial community and specializes in producing the renieramcyins.

**Discussion**

The *Halichondra–Ca. E. renieramycinifaciens* system described here differs from other examples of defensive symbiosis in marine sponges in several aspects. First, unlike *H. spongidae* and *H. vanosui* symbioses, *Ca. E. renieramycinifaciens* is an intracellular symbiont of a specific clade of marine sponges, where it dominates the bacterial community and specializes in producing the renieramcyins.

**A specific symbiosis between Ca. E. renieramycinifaciens and a clade of Halichondra sponges.** To gain insights into the distribution of *Ca. E. renieramycinifaciens* in the ocean, we compared the bacterial community composition of our four *Halichondra* sponges with that of 1,435 marine samples belonging to 1,194 sponges, 37 sediment metagenomes, 195 seawater metagenomes and 9 unclassified metagenomes from a Global Sponge Microbiome (GSM) analysis using high-throughput 16S rRNA gene sequencing (see Methods). The 1,194 GSM sponge metagenomes belonged to 96 types of sponges, 89 of which have genus-level or species-level assignments, whereas 7 are unclassified (Supplementary Table 5). After data processing using the Qiime pipeline, we searched all samples for operational taxonomic units (OTUs) matching *Ca. E. renieramycinifaciens* and mapped their distribution to a phylogenetic tree of the host sponges that we constructed based on their 18S rRNA gene sequences. Overall, we discovered OTUs matching *Ca. E. renieramycinifaciens* exclusively in sponge samples and not in the seawater or sediment samples included, supporting its obligate symbiotic lifestyle. Importantly, sequences matching *Ca. E. renieramycinifaciens* exist specifically in one clade of *Halichondra* sponges, namely, the four from this study and seven ‘blue Halichondras’ from the Pacific Ocean (Fig. 6), and is absent in all other sponges including *Halichondra* individuals that belong to other clades. Finally, when present, the relative abundance of *Ca. E. renieramycinifaciens* in *Halichondra* sponges is unusually high, ranging from 10% to 75% of the total 16S rRNA gene sequences in a given sample (Fig. 6 and Supplementary Fig. 9). These results establish *Ca. E. renieramycinifaciens* as an intracellular symbiont of a specific clade of marine sponges, where it dominates the bacterial community and specializes in producing the renieramcyins.
Ca. Entotheonella sp., Ca. E. renieramycinifaciens does not reside in the mesohyl or pinacoderm of the sponge, but lives instead in specialized sponge chemobacteriocytes. Second, Ca. E. renieramycinifaciens harbours an extremely reduced genome that is incapable of supporting a free-living state (by contrast, Ca. Entotheonella sp. harbours large, almost intact genomes)\(^{10,39,40}\). Last, it is predominantly a one-molecule system, in which ren is the only (or one of three) small-molecule BGC recovered from several deeply sequenced metagenomes, and the renieramycins are the major molecules observed (by contrast, dozens of BGCs and molecules are usually recovered from Ca. Entotheonella-harbouring sponges)\(^{10,39,40}\). However, several of these features are paralleled in cases of defensive symbioses of other host organisms\(^{41}\). Small-molecule-producing symbionts with reduced genomes and a known or presumed intracellular lifestyle have been reported from several marine ascidians, including the euteiniacidin producer Ca. Endotheinascidia frumentensis\(^{42,43}\). It is interesting that two distant bacteria, Ca. E. frumentensis and Ca. E. renieramycinifaciens (83% 16S rDNA gene sequence identity), which live in an intracellular symbiotic lifestyle with two very distant marine animals (an ascidian and a sponge), base their symbioses on the production of very similar defensive molecules (the euteiniacids and the renieramycins).

In insects, defensive symbionts with extremely reduced or evolutionarily degenerated genomes are relatively common\(^{44}\). For example, specialized cells of the psyllid Diaphorina citri contain an intracellular symbiont, Ca. Ploffiella armature, which harbours an extremely reduced genome (~460kb) and produces the defensive molecule diaphorin\(^{45}\). Similarly, Lagria villosa beetles harbour a genome-reduced symbiont, Burkholderia gladioli Lv-StB, both in female accessory glands and on eggs, which produces the defensive molecule lagriamide\(^{46}\). It is remarkable that widely different symbionts, hosts and molecules are involved in otherwise very similar defensive strategies.

Bacteriocytes have been previously observed in marine sponges, but the identity and function of the bacterial symbionts residing in them have been rarely studied\(^{47,48}\). The Ca. E. renieramycinifaciens–Haliclona system described here represents the second case in which intracellular sponge symbionts have been identified\(^{49}\) and one in which the symbiont genome has been sequenced. Hemimycale sponges harbour intracellular bacterial symbionts in specialized spherical cells termed calcibacteriocytes. In these cells, calcibacteria are cultivated in sponge vacuoles where the pH promotes the nucleation of calcium carbonate in bacterial membranes, which are later exported to deposit on the sponge exoskeleton\(^{50}\). Similarly, Ca. E. renieramycinifaciens are cultivated in Haliclona sponge chemobacteriocytes, where they produce the renieramycins. Altogether, it is remarkable that sponges—which have a very limited number of cell types—source essential structural and defensive roles to symbionts that are cultivated in specialized cells (calcibacteriocytes and chemobacteriocytes). Taking place in the oldest-living metazoans, these examples represent an ancestral view of microorganism–host interactions and highlight the importance of uncovering more cases of intracellular symbioses in marine sponges.

### Methods

**Sponge sample processing and storage.** Sponge samples were collected by scuba diving from different locations in the Pacific Ocean: Ren-PNG-07113: 5° 17.5’ S, 150° 06.1’ E; Ren-Pal-02: 7° 30’ N, 134° 30’ E. For Palau and Papua New Guinea samples, freshly collected samples were processed in the field for chemistry and DNA work. For chemistry work, a portion of each sponge sample was cut and frozen immediately. For DNA work, a portion was stored in RNAlater solution (Thermo Fisher Scientific). Ren-Bali-16-03 was freshly collected in Bali from a commercial mariculture farm (Quality Marine), shipped alive to the laboratory and processed immediately on arrival. For chemistry and DNA, Ren-Bali-16-03 was processed as described above; for microscopy, small portions of Ren-Bali-16-03 were first fixed in paraformaldehyde (4%), sequentially dehydrated in ethanol (30%, 50% and 70%) and finally stored in 70% ethanol at ~20°C until use. For flow cytometry, see below.

**DNA extraction and sequencing.** Metagenomic DNA was extracted from each RNAlater-preserved sponge sample using the Genomic Tip Kit (Qiagen) with some modifications. The sponge tissue (~1 cm³) was flash frozen in liquid nitrogen and homogenized using a sterile pestle in a conical tube. Buffer B1 (3.5 ml) containing RNase A (0.2 mg ml⁻¹) was then added and further homogenization was done. The solution was then treated with protease K (0.5 mg ml⁻¹) and lysozyme (2.5 mg ml⁻¹) for 5 h at 37°C while shaking. The Genomic Tip Kit Low Capacity DNA Lib Prep kit (Qiagen) was added and the solution was further incubated at 50°C for 30 min. Samples were then centrifuged and the supernatant was treated according to the Genomic Tip protocol (Qiagen). Metagenomic DNA was mechanically sheared to an average size of ~500 bp, and Illumina sequencing libraries were prepared using the Apollo Metagenomic Library Prep System (DepthSeq). The DNA libraries were sequenced on an Illumina HiSeq 2500 Rapid Flow cell as paired-end 2 × 157bp reads for Ren-PNG-07113, Ren-Pal-02 and Ren-Bali-16-03, and 2 × 141 bp for Ren-PNG-07060 (Supplementary Table 1).

**Metagenomic analysis, genome assembly and annotation.** Raw Illumina reads were filtered as follows using PRINSEQ\(^{51}\). Reads below an average quality score of ≥30 and with more than 2% undetermined (N) bases were discarded, bases with quality scores below 30 on either end of a read were trimmed and trimmed reads shorter than half of the original read length were discarded. Filtered reads (pairs and singletons) were assembled using SPAdes with default parameters\(^{52}\).

The four chromosomes and four plasmids were circularized by either mapping corresponding paired-end filtered reads to the edges of the SPAdes scaffolds using BLASTn or by simply aligning the scaffold edges and then repeating the assembly in Geneious with manual inspection\(^1\). The general overview of all functions in the Ca. E. renieramycinifaciens and Coxiella HT99 genomes was obtained by comparison of all the TIGRfam categories in the IMG. Analysis of individual pathways was done by examining the Cluster of Orthologous Group (COG), TIGRfam and KEGG pathways for the presence or absence of genes in each pathway. The circular map was downloaded from the IMG with standard COG categories as annotated\(^{47}\).

**Identification of common scaffolds between metagenomes.** The scaffolds from the SPAdes assembly of Ren-Bali-16-03 were filtered to a minimum length of 5,000 bp. To identify the scaffolds that are in common between Ren-Bali-16-03 and the other three metagenomes, filtered reads from the three metagenomes were mapped to Ren-Bali-16-03 scaffolds of >5,000 bp using Bowtie 2 (end-to-end alignment, fast mode, with the minimum alignment score set as L = −0.6, −0.3)\(^{53}\). The breadth coverage of each scaffold (the percentage of the scaffold length covered by reads) was then calculated using SAMtools scripts, and scaffolds with <90% breadth coverage were discarded\(^{53}\). Scaffolds that were covered by reads from all three metagenomes using the above cut-off were counted as common to all metagenomes and then taxonomically assigned using BLASTx against the non-redundant (nr) protein database on the NCBI.

**16S amplicon sequencing and meta-analysis.** The V4 region of the 16S rDNA gene (~250 bp) was PCR amplified and used to construct Illumina sequencing libraries following the previously published design and primers\(^{54}\). Libraries were pooled, sequenced on an Illumina HiSeq 2500 Rapid Flow cell as paired-end reads of 2 × 175 bp, along with 8-bp index reads, following the manufacturer’s protocol (Illumina). Raw reads were demultiplexed based on their index reads, and overlapping forward and reverse reads were merged using FLASH with a minimum overlap of 80, a maximum overlap of 100 and a maximum mismatch density of 0.2 (ref. \(^{55}\)). The resulting FASTQ sequences were preprocessed in Qita (qita.ucsd.edu) using split_libraries.py. The resulting FASTA sequences were trimmed to the first 100 nucleotides and the resulting file was downloaded for downstream processing in Qiime (qiime.org)\(^{56}\). To compare the bacterial composition of Haliclona sponges with the existing sponge microbiome data, we used a published data set that we refer to here as the GSM\(^{48}\). The demultiplexed, trimmed, FASTA sequences from the GSM were downloaded from Qita (study 10346). The GSM and Haliclona sequences were then combined and submitted to subsequence open-reference OTU picking using pick-open-reference-otus.py in Qiime using default parameters, except for enabled reverse-strand matching and the number of parallel jobs. OTUs were defined at 97% identity. The resulting OTU table was filtered to a 0.005% minimum relative abundance and then split according to sample type (sponge metagenomes, seawater metagenomes) and sediment metagenomes. Jaccard similarity was calculated from the total number of OTUs in each host. A phylum-level summary of the OTU table with Ca. E. renieramycinifaciens OTU as a group was constructed for mapping with the sponge phylogeny (described below).
Mapping of microbial composition to sponge phylogeny. To construct the phylogenetic tree of host sponges, the 18S rRNA gene sequences were amplified from the metagenomic DNA of the Haliclona sponges using 18S rRNA gene universal primers 18S-A1 (5′-ACCTGGTTGATCCTGCGGAC-3′) and 18S-564R (5′-GGCACCAAGAAGTCTTCC-3′). The following thermocycler programme was used: initial denaturation, 1 min at 98 °C, 34 cycles of 98 °C for 10 s, 65 °C for 30 s, 72 °C for 30 s; and final extension of 5 min at 72 °C. Phusion high-fidelity polymerase (New England Biolabs) was used. The amplification products were gel purified and sequenced with direct primers. E. renieramyci faciens was cloned into a pET28 vector under the control of the E. coli promoter and transformed into Stellar cells (Takara Bio). Heterologous expression of E. renieramyci faciens was performed at 46 °C for 2–3 h in pre-hybridized buffers with the same hybridization buffers. The samples were then mounted on slides and imaged. The optimised formamide concentration, 35%, as determined from this experiment, was used for the hybridization on sponge sections using the same procedure as above except for the additional counterstaining with DAPI before FISH. Probe CE75 and the corresponding mismatch probe were labelled with 6-FAM and the universal eubacterial probes were labelled with Cy3. Imaging was performed on a Leica SP5 laser scanning confocal microscope, using 405-nm, 480-nm and 581-nm lasers and a x63 magnification oil objective.

TEM sample preparation and imaging. A small cube (~0.5 cm³) was cut from a paraformaldehyde-fixed sample of Ren-Bali-16-03 and placed into a 2% osmium tetroxide solution (1 ml). Staining was done overnight on a rotator, after which the sample was washed five times with ddH₂O and then stained with 2% uranyl acetate for another 12 h. After washing with ddH₂O, the sample was then dehydrated sequentially in 30%, 50%, 70%, 85% and 100% ethanol for 1 h at each step. The dehydrated sample was embedded in 1:1 LR white resin (Electron Microscopy Sciences)/ethanol overnight on a rotator and then transferred to 100% LR white resin for stationary embedding for 48 h. Polymerization was performed on a hot block at 65 °C overnight. Ultrathin sections were observed on a Talos F200X Scanning/Transmission Electron Microscope.

LCM of sponge chemobacteriocines. A small cube (~0.5 cm³) of RNA Laser- preserved sample was washed in sterile water three times to reduce salt crystals. The tissue was then gently cut on a 70-μm filter, which was then washed with sterile water (500 μl). The filtrate was centrifuged at 500g for 5 min at 4 °C to pellet small sponge cells and particles. The resulting pellet was then suspended in water (100 μl), and a 10% dilution (50 μl) was spread on a LCM PEN membrane (Thermo Fisher Scientific) and allowed to air dry. Chemobacteriocines were identified as distinct round particles with a diameter of 15–25 μm and isolated by laser microdissection on an MM Cell Cut LCM system (mmi). One-hundred bacteriocines were collected and 50 background membrane cuts were further collected as a negative control. DNA was extracted from both the cells and the negative controls using the Masterpure complete DNA extraction kit following the manufacturer’s protocol (Epicenter) and used for Illumina library preparation and metagenomic sequencing as described above (Supplementary Table 1). BLASTn (e-value cut-off of 1 × 10⁻³⁰) was used to assign metagenomic reads to the Ce. renieramyci faciens chromosome, p-reno and sponge mitochondrion.

Chemical extraction and analysis. Frozen sponge samples (~2 cm³) were cut and extracted twice with ethyl acetate (20 ml) and twice with methanol (20 ml). The extracts were combined, dried under vacuum and analysed by HPLC–HR-IC/ MS/MS on an Agilent QTOF instrument (Agilent Technologies). HPLC elution was done using the following gradient: 0–100% A, 0–25 min; 100% A, 25–30 min; and 100% B, 30–35 min (A, 0.1% formic acid in water; B, 0.1% formic acid in acetonitrile (v/v)). Purification renieramycinin E, the frozen Ren-Bali-16-03 sponge (~5 g) was extracted as described above. The dried extract was subjected to fractionation using a semi-preparative HPLC column (Agilent Poroshell 120 EC-C18, 9.4×250 mm, 4 μm) on an Agilent 1260 Infinity HPLC system (Agilent Technologies). The following gradient was used for fractionation: 0–100% B, 0–20 min; 100–0.5% B, 20–25 min. The flow rate for fractionation was 1.5 ml/min. Fractions were analysed immediately using HPLC–MS, and the fraction containing the major peak and the m/z corresponding to renieramycin E (549.22 m/z (M + H²O)⁺) was dried and submitted for proton NMR. NMR spectra were obtained on an Avance III, 500 MHz (Bruker) in CD3OD. We successfully obtained the proton NMR of a semi-pure renieramycin E sample that showed distinctive renieramycin signatures (for example, O-methyl groups (3.91 ppm) and angelic acid methyl groups (1.66 ppm, 1.47 ppm) (Supplementary Fig. 2), but owing to the typical instability of renieramycins and THQs in general, further purification attempts yielded degraded products.

Heterologous expression of renfl and ren in E. coli. The C-methyltransferase gene, renf, was amplified from the Ren-Bali-16-03 metagenomic DNA using primers (forward 5′-GGTTACCGTTGAAGAAATATTTGTGTTAGAAAGAAAGT-3′, reverse 5′-GCTAGTTGGATGTATCGTTCTAATTGTCATTCCGC-3′), and cloned into the pGFP-UV backbone by isothermal assembly using the In-Fusion cloning kit (Takara Bio), under the control of the lac promoter. For heterologous expression, a pGFP-renf vector was transformed into an E. coli cell line (Takara Bio). The pGFP-UV vector was used as a negative control. Colonies were picked from the transformations and grown overnight at 300 °C at 200 r.p.m. Seed cultures (100 μl) were inoculated in LB broth (100 ml) that was supplemented with carbenicillin (100 μg ml⁻¹) and L-tyrosine (5 mM) (Thermo Fisher Scientific). Expression cultures were grown at 300 °C at 200 r.p.m. in 5-ml beakers. After expression, the E. coli cells were harvested by centrifugation. To obtain the extracts, be pre-equilibrated (5 ml) Dianon HP20 absorbent resin (Millipore-Sigma) was added to the supernatant and bound by gently shaking at 150 r.p.m. at room temperature for 1 h. The resin...
was then filtered and washed with three volumes of dH2O, and finally eluted with methanol (20 mL). The methanolic extracts were dried and analysed using HPLC-HR-MS as described above. Authentic standards of 3-methyl-t-yrosine and O-methyl-t-yrosine (Millipore-Sigma) were purchased and compared to the extracts of renB expression. To express the entire ren BGC, three overlapping fragments of ~9 kb were amplified from Ren-Pal-02 metagenomic DNA using the following primers: natren_514F: 5'-AATTGATGAACGAGCTGCGC and natren_5954R: 5'-CACAGTATCTATACTGCGT; natren_8409F: 5'-CAACCTGCTCTATACGCTG and natren_17633R: 5'-CTGGTATCAAGGTGTGGCC and natren_16615F: 5'-TGCTGATGCACGACAAATCTC and natren_25404R: 5'-CAGCTAAATCCTCCATCCCA. Transformation-associated recombination in Saccharomyces cerevisiae was performed using established methods to assemble all three fragments (ren) into a plasmid designed based on the pGGP-UV backbone and the yeast elements from pLLX13 (refs. 64,65). This construct was expressed in E. coli EP300 cells using the native ren promoters in LB medium for 5 d at 30°C and 200 r.p.m. To co-express the ren gene cluster with a phosphopantetheinytransferase and an MBH-like protein identified in Ca. E. renieramycini, genes encoding them were amplified from the metagenomic DNA of Ren-Pal-02. For the phosphopantetheinytransferase, primers forward 5’-AATTTCCACACGAGAAAACGCTATGAACAACTTGCAGGGA and reverse 5’-ACAGGATATACTGACAAATTTACTTGGAGGTTCC were used and the fragment was cloned into the pSTV28 backbone (Takara Bio) downstream of the lac promoter. The gene encoding an MBH-like protein was amplified using primers forward 5’-GCCGTACAGTGGATATTCGATACCACTGCGTCTGCCAGGGC and reverse 5’-CGCGCTAGCTGATGAGAATATGTTTAGGATGTTGCAGGTTG cloned and integrated into the pGGPUV plasmid after the lac promoter (Takara Bio). The fragment including the lac promoter and the gene encoding MBH from the pGGPUV-mbth construct was subsequently transferred to the pSTV28-pptase construct to yield the pSTV28-pptase-mbth plasmid. Plasmids were constructed using Gibson isothermal assembly. The resulting pSTV28-pptase-mbth plasmid was transformed into E. coli EP300 cells containing the ren pathway for expression. Expression was performed in LB medium for 5 d at 30°C and 200 r.p.m., and extracts were obtained and analysed as described above.

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

Data availability. The data that support the findings of this study are available from the corresponding author on request. The Ca. E. renieramycini genomes have been deposited to the IMG (Joint Genome Institute, Department of Energy) public repository, under IMG submission IDs 151197, 151198, 119799 and 119800.

Received: 31 July 2018; Accepted: 18 February 2019; Published online: 1 April 2019

References
1. Thomas, T. et al. Diversity, structure and convergent evolution of the global sponge microbiome. Nat. Commun. 7, 11870 (2016).
2. Hentschel, U., Piel, J., Degnan, S. M. & Taylor, M. W. Genomic insights into the marine sponge microbiome. Nat. Rev. Microbiol. 10, 641–654 (2012).
3. Nguyen, M. T., Liu, M. & Thomas, T. Ankyrin-repeat proteins from sponge E. renieramycini. J. Proteome Res. 13, 1095–1104 (2014).
4. Burgsdorf, I. et al. Lifestyle evolution in cyanobacterial symbionts of sponges. Proc. Natl Acad. Sci. USA 109, 22502–22507 (2012).
5. Tsukihara, K. et al. Novel bacterial genomic islands associated with the marine sponge Reniera renieramycini. J. Bacteriol. 196, 3754–3762 (2014).
6. Shackel, J., Hulsen, R. & Bresch, A. Novel gene clusters for the biosynthesis of the antibiotic saframycin Mx1 encoding a peptide synthetase. Microbiology 141, 1793–1803 (1995).
7. Weber, T. et al. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res. 43, W237–W243 (2015).
8. Woodhouse, J. N., Fan, L., Brown, M. V., Thomas, T. & Neilan, B. A. Deep sequencing of non-ribosomal peptide synthetases and polyketide synthases from the microbiomes of Australian marine sponges. ISME J. 10, 1842–1851 (2016).
9. Fu, C.-Y. Biosynthesis of 3-hydroxy-5-methyl-0-methylrhizosine in the saframycin/safracin biosynthetic pathway. J. Microbiol. Biotechnol. 19, 439–446 (2009).
10. Velasco, A. et al. Molecular characterization of the safracin biosynthetic pathway from Pseudomonas fluorescens A2-2: designing new cytotoxic compounds. Microbiol. 56, 144–154 (2005).
11. Amann, R. I. et al. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Appl. Environ. Microbiol. 56, 1919–1925 (1990).
12. Daims, H., Bruhl, A., Amann, R., Schleifer, K. H. & Wagner, M. The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. Syst. Appl. Microbiol. 22, 434–444 (1999).
13. Webster, N. et al. Same, same but different: symbiotic bacterial associations in GBR sponges. Front. Microbiol. 3, 444 (2012).
14. Checco, G. et al. Marine cyanobacteria: an integrated genomics resource for synthesis and metabolic diversity. Proc. Natl Acad. Sci. USA 111, 289–294 (2014).
15. Daims, H., Schleifer, K. H. & Wagner, M. The domain-specific probe EUB338 is insufficient for the detection of all mesophilic bacteria: evaluation of a more comprehensive probe set. Syst. Appl. Microbiol. 22, 434–444 (1999).
16. Webster, N. et al. Same, same but different: symbiotic bacterial associations in GBR sponges. Front. Microbiol. 3, 444 (2012).
17. Checco, G. et al. Microbial Integrated Genomics resource for synthesis and metabolic diversity. Proc. Natl Acad. Sci. USA 111, 289–294 (2014).
18. Caporaso, J. G. et al. QIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, 335–340 (2010).
19. Lackner, G., Peters, E. E., Helfrich, E. J. & Piel, J. Insights into the lifestyle of uncultured bacterial natural product factories associated with marine sponges. Proc. Natl Acad. Sci. USA 114, E34–E356 (2017).
40. Mori, T. et al. Single-bacterial genomics validates rich and varied specialized metabolism of uncultivated Entothoeonella sponge symbionts. Proc. Natl Acad. Sci. USA 115, 1718–1723 (2018).

41. Florez, L. V., Biedermann, P. H., Engl, T. & Kaltenpoth, M. Defensive symbioses of animals with prokaryotic and eukaryotic microorganisms. Nat. Prod. Rep. 32, 904–936 (2015).

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

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

44. Engl, T. et al. Evolutionary stability of antibiotic protection in a defensive symbiosis. Proc. Natl Acad. Sci. USA 115, E2020–E2029 (2018).

45. Nakabachi, A. et al. Defensive bacteriome symbiont with a drastically reduced genome. Curr. Biol. 23, 1478–1484 (2013).

46. Maldonado, M. Intergenerational transmission of symbiotic bacteria in oviparous and viviparous demosponges, with emphasis on intracytoplasmically-compartmented bacterial types. J. Mar. Biol. Assoc. UK 87, 1701–1713 (2007).

47. Vacelet, J. & Donadey, C. Electron microscope study of the association between some sponges and bacteria. J. Exp. Mar. Biol. Ecol. 30, 301–314 (1977).

48. Uriz, M. J., Agell, G., Blanquer, A., Turon, X. & Casamayor, E. O. Endosymbiotic calcifying bacteria: a new cue to the origin of calcification in metazoa? Evolution 66, 2993–2999 (2012).

49. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27, 863–864 (2011).

50. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477 (2012).

51. Kearse, M. et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649 (2012).

52. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012).

53. Li, H. et al. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).

54. Caporaso, J. G. et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 6, 1621–1624 (2012).

55. Magoc, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27, 2957–2963 (2011).

56. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797 (2004).

57. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS ONE 5, e9490 (2010).

58. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44, W242–W245 (2016).

59. Albertsen, M. et al. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat. Biotechnol. 31, 533–538 (2013).

60. Srivastava, M. et al. The Amphimedon queenslandica genome and the evolution of animal complexity. Nature 466, 720–726 (2010).

61. Wickham, H. ggplot2: Elegant Graphics for Data Analysis (Springer, 2009).

62. Ludwig, W. et al. ARB: a software environment for sequence data. Nucleic Acids Res. 32, 1363–1371 (2004).

63. Schramm, A., Fuchs, B. M., Nielsen, J. L., Tonolla, M. & Stahl, D. A. Fluorescence in situ hybridization of 16S rRNA gene clones (clone-FISH) for probe validation and screening of clone libraries. Environ. Microbiol. 4, 713–720 (2002).

64. Wolfgang, M. C. et al. Conservation of genome content and virulence determinants among clinical and environmental isolates of Pseudomonas aeruginosa. Proc. Natl Acad. Sci. USA 100, 8484–8489 (2003).

65. Yamanaka, K. et al. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. Proc. Natl Acad. Sci. USA 111, 1957–1962 (2014).

Acknowledgements

We thank E. W. Schmidt, M. K. Harper-Ireland and C. Ireland at the University of Utah for providing samples Ren-PNG-07060, Ren-PNG-07113 and Ren-Pal-02, and the Republic of Palau, Papua New Guinea and the Republic of Indonesia as original sources for the sponge samples studied here. We thank M. K. Harper-Ireland and C. Ireland at the University of Utah for the underwater photograph of the Halichondra sponge shown in Fig. 1. We are grateful to C. DeCoite and the Molecular Biology Flow Cytometry Resource Facility (partially supported by the Cancer Institute of New Jersey Cancer Center Support Grant P30CA072720) for assistance with flow cytometry; P. Shao and the Molecular Biology Electron Microscopy Core Facility for assistance with TEM; G. Laevsky, the Molecular Biology Confocal Microscopy Core Facility (a Nikon Center of Excellence) and I. Zan for assistance with FISH and microscopy experiments; G. Hrebikova and A. Ploss for assistance with LCM; W. Wang and the Lewis Sigler Institute Sequencing Core Facility for assistance with high-throughput sequencing; M. Cahn for assistance with metagenomic data analysis; S. Chatterjee for general assistance; Y. Sugimoto and P. Chankhamjem for assistance with NMR and HPLC–HR-MS; and the rest of the Donia lab for useful discussions. We also thank the anonymous Nature Microbiology reviewer who suggested the name Ca. e. renieramycinifaciens for the symbiont discovered in this study. Funding for this project has been provided by Princeton University, and M.S.D. is funded by an NIH Director’s New Innovator Award (ID: 1DP2AI124441).

Author contributions

M.D.T.-M. and M.S.D. designed the study; M.D.T.-M., I.N.B. and M.S.D. performed the experiments, analysed the data and wrote the manuscript.

Competing interests

M.S.D. is a member of the Scientific Advisory Board for Deepbiome Therapeutics and a consultant for Flagship Pioneering.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-019-0415-8.

Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.S.D.

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

© The Author(s), under exclusive licence to Springer Nature Limited 2019
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement

- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly

- The statistical test(s) used AND whether they are one- or two-sided

- Only common tests should be described solely by name; describe more complex techniques in the Methods section.

- A description of all covariates tested

- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons

- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)

- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted

- Give P values as exact values whenever suitable.

- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings

- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes

- Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

- Clearly defined error bars

- State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection

FACSDiVa software (BD Biosciences, San Jose, CA) was used for flow cytometry data acquisition.

Data analysis

SPAdes 3, antiSMASH 3, Qiime 1, IMG (web tool), Qiita (web tool), PRINSEQ (Version 0.20.4B lite), Bowtie2 (Version 2.2.9), FLASH (Version v1.2.11), MUSCLE (Version 3.8.425), Fasttree-ML (Version 2.1.10), iTOL (Version 4), Rstudio (Version 1.0.143), ARB (Version 6.0.4), MEGA (Version 7). Any deviations from the defaults parameters are explained in the Methods section.

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

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data availability statement is in line 931-934, and again here: Ca. E. renier genomes have been deposited to the Integrated Microbial Genomes (Joint Genome Institute, Department of Energy) public repository, under IMG submission IDs 151197, 151198, 119799, 119800.

## Field-specific reporting

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

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

### Ecological, evolutionary & environmental sciences study design

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

#### Study description

This study describes the discovery of the biosynthetic origin and the microbial producer of renieramycin compounds from marine sponges using combined metagenomics, chemistry, heterologous expression, 16S analysis, flow cytometry, laser capture microdissection, and Fluorescence in-situ hybridization techniques. The exact details, number of experiments performed for each technique, and replicates employed are explained in the manuscript.

#### Research sample

Samples of blue/purple Haliclona sponges were collected from different years (2002, 2007, and 2016) and different locations (Palau, Papua New Guinea, and Bali). We focused on blue/purple Haliclona sponges because they are known historically to contain the renieramycins. We also compared microbiome data from a previous study to ours: Global Sponge Microbiome analysis (GSM), Qiita study 10346.

#### Sampling strategy

Sampling from different years and locations allowed the identification of common, host-specific, bacterial symbionts and biosynthetic gene clusters in blue/purple Haliclona sponges that are associated with the production of renieramycins. Our sampling strategy relied mainly on identifying different Haliclona sponge individuals that contain renieramycins. We were only able to identify 4 such individuals, and all of them were characterized in this study. Since the goal of the study is to identify the bacterial symbiont and biosynthetic pathways responsible for renieramycin production, there was no need to include additional samples or perform statistical analyses for sample size calculations.

#### Data collection

Metagenomic sequencing was performed at the Lewis Sigler Institute sequencing core facility at Princeton University, and analyzed by the authors of the manuscript. Chemical analysis of the collected samples was performed by the authors of the manuscript on Mass Spectrometry instruments available in the Donia laboratory at Princeton University.

#### Timing and spatial scale

Samples of blue/purple Haliclona sponges collected from different years (2002, 2007, and 2016) and different locations (Palau, Papua New Guinea, and Bali). The rationale was to study how stable symbiont-derived renieramycin production is throughout time and space, given our limited access to samples and collection sites.

#### Data exclusions

No data was excluded.

#### Reproducibility

1. We analyzed four independent samples of renieramycin-containing Haliclona sponges to ensure the reproducibility of our finding, which is connecting the ren BGC to renieramycin production. 2. To prove that the ren gene cluster found in sponge metagenomes encodes for the renieramycins, we performed heterologous expression of the C-methyltransferase that we predicted to be responsible for converting the tyrosine substrate into 3-methyltyrosine (renB). We expressed this enzyme in E.coli in the presence of added tyrosine and analyzed the data by LC/HRMS comparison to standard compounds. We performed the expression experiments 3 times, with triplicated samples each time for the control (Ecoli transformed with empty vector) and the expressions (E. coli transformed with cloned renB). 3. We performed Fluorescence in-situ hybridization on several sections of Ren-Bali-16-03, and always obtained the same localization results. All of the attempts mentioned above were successful and ensured the reproducibility of our findings.

#### Randomization

We did not need to randomize the samples, because they were all selected in the first place to contain the renieramycins.

#### Blinding

We did not need to blind the identity of the samples, because they were all selected in the first place to contain the renieramycins.

### Did the study involve field work?

- Yes
- No
Field work, collection and transport

Field conditions
For the Palau sample Ren-Pal-02, sample was collected in a murky river mouth region. For the two Papua New Guinea samples, they were collected on a tropical reef face, clear water.

Location
Sponge samples were collected by SCUBA diving from different locations in the Pacific Ocean: Ren-PNG-07060, 5° 17' S. 150° 06' E.; Ren-PNG-07113, 5° 17.5' S. 150° 06.1' E; and Ren-Pal-02: 7° 30' N. 134° 30' E. Ren-Bali-16-03 was freshly collected in Bali from a commercial mariculture farm (Quality Marine, USA).

Access and import/export
Papua New Guinea samples were collected according to Permit Number 011299 for 2007 for exporting wild life, issued by the Department of Environment and Conservation at Papua New Guinea, between Dr. Chris Ireland and Dr. Louis Barrows at the College of Pharmacy, University of Utah and Dr. Lohi Matainaho at the School of Medicine and Health Sciences, Papua New Guinea. The Palau sample was collected with a Permit arranged between the late John Faulkner at the Scripps Institute of Oceanography, University of California, San Diego, and the Coral Reef Research Foundation in Palau. Ren-Bali-16-03 was obtained commercially through Quality Marine, USA.

Disturbance
Very small portions of the sponges were collected and only minimal disturbance was caused to the environment during the collection.

Reporting for specific materials, systems and methods

Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
|     | □ Unique biological materials |
| □   | Antibodies |
| □   | Eukaryotic cell lines |
| □   | Palaeontology |
| □   | Animals and other organisms |
| □   | Human research participants |

Methods

| n/a | Involved in the study |
|-----|-----------------------|
| □   | ChIP-seq |
| □   | Flow cytometry |
| □   | MRI-based neuroimaging |

Animals and other organisms

Policy information about studies involving animals: ARRIVE guidelines recommended for reporting animal research

Laboratory animals
The study did not involve laboratory animals

Wild animals
The study did not involve wild animals

Field-collected samples
Sponge samples were collected by SCUBA diving. Freshly collected samples were processed in the field for chemistry and DNA work as follows. For chemistry work, a portion of each sponge sample was cut and frozen immediately at -20 °C. For DNA work, a portion was stored in RNAlater solution (ThermoFisher Scientific, USA), then frozen at -20 °C. Ren-Bali-16-03 was shipped alive to the laboratory, and processed immediately upon arrival for chemistry and DNA as described earlier in this section. For microscopy, small portions of Ren-Bali-16-03 were first fixed in paraformaldehyde (4%), sequentially dehydrated in ethanol (30 %, 50%, and 70 %), and finally stored in 70% ethanol at -20 °C until use.

Flow Cytometry

Plots
Confirm that:
- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation
A 1 cm³ of fresh sponge sample was gently homogenized using a sterile tube and pestle and was filtered through a 70 micron filter and washed with sterile artificial sea water. The filtrate was then diluted 5 times and used directly for flow cytometry.
Instrument FACSVantage SE w/DiVa cell sorter (BD Biosciences, San Jose, CA USA)

Software FACSDiVa software (BD Biosciences, San Jose, CA) was used for both data acquisition and analysis.

Cell population abundance Partitions of cells obtained from cell sorting were subjected directly to metagenomic sequencing analysis and 16S rRNA gene amplicon sequencing analysis.

Gating strategy Particles were observed on the forward scatter (FSC) and side scatter (SSC) parameters, representing increasing 488 nm laser light scatter due to particle size and granularity, respectively. Eight gates were created along the FSC axis and particles from each were collected into separate tubes, resulting in partitions with increasing particle size composition.

.Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.