

RESEARCH ARTICLE SUMMARY

SYMBIOSIS

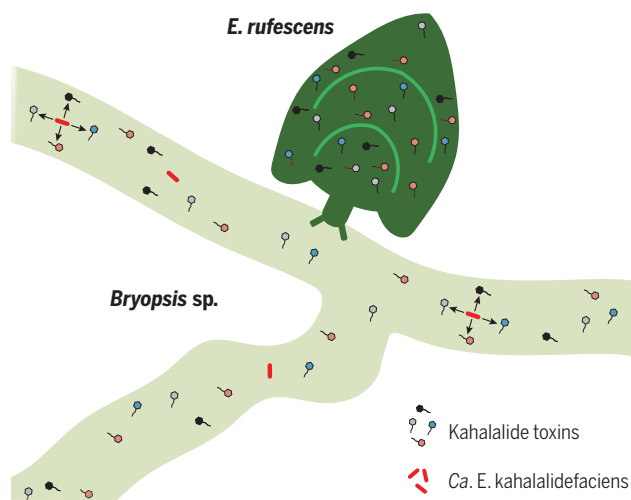
A microbial factory for defensive kahalalides in a tripartite marine symbiosis

Jindong Zan*, Zhiyuan Li*, Ma. Diarey Tianero, Jeanette Davis, Russell T. Hill, Mohamed S. Donia†

INTRODUCTION: Chemical defense strategies, in which organisms use toxic molecules for protection against pathogens or predators, are widespread in the marine environment. In some cases, the same defensive molecules are shared by taxonomically distant organisms, raising questions about their molecular origin. The actual source of these molecules may be the organism itself, as observed in marine algae; a microbial symbiont, as commonly seen in marine sponges and tunicates; or diet, as in several marine mollusks. Elucidating the molecular basis of toxin production in chemically defended organisms is important for a complete understanding of their ecological interactions.

RATIONALE: In this work, we studied toxin production in the Hawaiian marine alga *Bryopsis* sp. and its predator, the mollusk *Elysia rufescens*. Both organisms are chemically defended against predators by a diverse library of lipopeptide toxins, the kahalalides, but the details of kahalalide production and diversification are unknown (see the figure). One of these molecules, kahalalide F, is a potent cytotoxin and has been evaluated clinically as an anticancer agent. The molecular structures of the kahalalides show several features of microbial biosynthesis: They are fatty acid–cyclic peptide hybrids with several D- and non-proteinogenic amino acids, thus motivating us to hypothesize that the kahalalides are produced by a bacterial or fungal symbiont of *Bryopsis* sp. or of both *Bryopsis* sp. and *E. rufescens*. We combined metagenomic, metatranscriptomic, and chemical analyses with microbial cultivation, fluorescence microscopy, and evolutionary genomics to determine the molecular bases of kahalalide production and evolution in this tripartite marine symbiosis.

RESULTS: Using metagenomic analyses, we discovered a bacterium—termed “*Candidatus* Endobryopsis kahalalidefaciens”—that has no cultured close relatives, and we show that it lives in symbiosis with the alga *Bryopsis* sp. (see the figure). Using fluorescence microscopy, bacterial cultivation, and comparative genomics, we show that “*Ca. E. kahalalidefaciens*” is an intracellular,



Chemical defense in a tripartite marine symbiosis. The bacterial symbiont “*Ca. E. kahalalidefaciens*,” which lives intracellularly in the marine alga *Bryopsis* sp., produces a diverse library of toxins (the kahalalides) that protect the host from predation. The mollusk *E. rufescens* sequesters the same toxins from its algal diet and employs them for its own defense.

obligate, genome-reduced symbiont that has lost essential functions for free living (e.g., amino acid biosynthesis). Despite this reduced metabolic capacity, 20% of the “*Ca. E. kahalalidefaciens*” genome encodes a diverse set of 20 nonribosomal peptide synthetase pathways. We link nine of these pathways to nine structurally diverse kahalalides (including kahalalide F, the main defensive toxin of both *Bryopsis* sp. and *E. rufescens*), which we then chemically identify in the same sample of *Bryopsis* sp.

None of the amino acid substrates that make up the kahalalides can be produced by “*Ca. E. kahalalidefaciens*” itself; therefore, these substances are mostly provided by the autotrophic *Bryopsis* sp., highlighting an unusual strategy of collaborative biosynthesis between a symbiotic bacterium and its host. Moreover, using metagenomic analysis and fluorescence microscopy, we show that “*Ca. E. kahalalidefaciens*” is not a symbiont of the

ON OUR WEBSITE

Read the full article at <http://dx.doi.org/10.1126/science.aaw6732>

mollusk *E. rufescens*, establishing chemical sequestration as the means by which this animal indirectly acquires bacterially produced kahalalides from its algal diet.

Detailed analysis of the “*Ca. E. kahalalidefaciens*” genome reveals a high level of plasticity and a distinctive model of diversifying evolution that is independent of horizontal gene transfer, consistent with the intracellular lifestyle of this symbiont. In this model, new non-ribosomal peptide synthetase pathways arise through duplication and divergence, accompanied by extensive interpathway recombination events. Finally, meta-transcriptomic analysis reveals that 26% of the transcriptional activity of “*Ca. E. kahalalidefaciens*” is dedicated to kahalalide biosynthesis and that biosynthetic pathways for different kahalalides vary widely in their expression levels, further emphasizing the importance of these molecules in maintaining a successful symbiosis.

CONCLUSION: In a chemically defended, tripartite marine symbiosis, we show that an obligate bacterial symbiont of a marine alga produces a library of defensive molecules that protect the host from predation, and that the same molecules are in turn hijacked by a predatory mollusk and used for its own defense. Living intracellularly in algal cells, the symbiont acts as a microbial factory for the biosynthesis of complex defensive molecules from simple host-derived substrates. Symbiont-derived production of defensive mol-

ecules in marine algae and indirect acquisition of microbial products by predatory mollusks may thus be important yet rarely studied phenomena in marine ecosystems. ■

The list of author affiliations is available in the full article online.

*These authors contributed equally to this work.

†Corresponding author. Email: donia@princeton.edu
Cite this article as J. Zan et al., *Science* 364, eaaw6732 (2019). DOI: 10.1126/science.aaw6732

RESEARCH ARTICLE

SYMBIOSIS

A microbial factory for defensive kahalalides in a tripartite marine symbiosis

Jindong Zan^{1*}, Zhiyuan Li^{2,3*}, Ma. Diarey Tianero¹, Jeanette Davis^{4,†},
Russell T. Hill⁴, Mohamed S. Donia^{1†}

Chemical defense against predators is widespread in natural ecosystems. Occasionally, taxonomically distant organisms share the same defense chemical. Here, we describe an unusual tripartite marine symbiosis, in which an intracellular bacterial symbiont (“*Candidatus* Endobryopsis kahalalidefaciens”) uses a diverse array of biosynthetic enzymes to convert simple substrates into a library of complex molecules (the kahalalides) for chemical defense of the host, the alga *Bryopsis* sp., against predation. The kahalalides are subsequently hijacked by a third partner, the herbivorous mollusk *Elysia rufescens*, and employed similarly for defense. “*Ca. E. kahalalidefaciens*” has lost many essential traits for free living and acts as a factory for kahalalide production. This interaction between a bacterium, an alga, and an animal highlights the importance of chemical defense in the evolution of complex symbioses.

One of the most efficient defense strategies in the ocean is chemical defense, in which organisms accumulate toxic small molecules that kill and/or deter predators (1). Because of the structural complexity of these defensive chemicals and their similarity to bacterial products, many have been proposed to originate from bacterial symbionts (2, 3). Indeed, defensive molecules of substantial toxicity have been genetically linked to specific bacterial symbionts of several marine animals (4). The concept of chemical defense can also extend beyond a single animal, with the same chemicals benefiting a network of partners. This phenomenon is common in marine mollusks that prey on chemically defended organisms and sequester the organisms’ toxic chemicals for their own defense (5). Although these two phenomena—production of defensive chemicals by bacterial symbionts and sequestration of defensive chemicals by predatory mollusks—have been separately described in marine systems, their integration in multipartite symbioses is rarely studied.

In this work, we investigated the marine alga *Bryopsis* sp. and its predator mollusk *Elysia*

rufescens (Fig. 1). *Bryopsis* sp. harbors a toxic lipopeptide molecule, kahalalide F (KF) (6), which was shown to be responsible for chemical defense against predators (7). Unlike other predators that are repelled by this toxin, *E. rufescens* not only feeds on KF-containing *Bryopsis* sp., but it also accumulates KF in its body at several times the concentration in *Bryopsis* sp. and employs KF for its own defense (7–9). Similarly to other species of *Elysia*, *E. rufescens* additionally maintains photosynthetically active algal chloroplasts in its digestive organs for several months, in a phenomenon known as kleptoplasty (10). Although KF has been isolated only from *Bryopsis* sp. and *E. rufescens*, it contains structural features that suggest a possible microbial origin, namely nonproteinogenic amino acids (e.g., ornithine, dehydrobutyrine, and several D-amino acids) and a fatty acid moiety (Fig. 1). These features motivated us to hypothesize that KF is produced by a cryptic third partner, such as a bacterial or fungal symbiont. To test this hypothesis and study the molecular and evolutionary details of toxin production in this unusual system, we employed a multidisciplinary approach that combines metagenomic, meta-transcriptomic, and chemical analyses; microbial cultivation; fluorescence microscopy; and evolutionary genomics.

KF production by a bacterial symbiont

We collected fresh *Bryopsis* sp. specimens from an algal bloom at the exact location where KF was originally reported (Black Point Bay, Honolulu, Hawaii; sample designated *Bryopsis*-2015) (Fig. 1A) and analyzed them using a variety of chemical and molecular techniques. First, we chem-

ically extracted *Bryopsis*-2015 as previously described (6), analyzed its resulting organic extract using high-performance liquid chromatography coupled with high-resolution tandem mass spectrometry (HPLC-HR-MS/MS), and confirmed that it harbors KF (fig. S1 and table S1). Second, to investigate the microbial community associated with *Bryopsis*-2015, we used high-throughput 16S ribosomal RNA (rRNA) gene amplicon sequencing (104,000 reads, on average, V4 region) on DNA isolated from multiple replicates of the same algal collection. Surprisingly, only three bacterial classes dominate the *Bryopsis*-2015 microbiome—Gammaproteobacteria (11.8% of the reads, on average), Alphaproteobacteria (9.3%), and Flavobacteriia (11.8%)—whereas most of the remaining reads (57.0%) map to the chloroplast-encoded 16S rRNA gene in *Bryopsis* sp. (Fig. 1C). A single sequence accounts for 74.1 to 92.0% of the fraction assigned to the class Flavobacteriia, making it the single most dominant species in *Bryopsis*-2015. Notably, four biological replicates collected on two consecutive days showed consistent microbiome profiles. These results reveal a relatively simple and uniform microbiome composition for KF-containing *Bryopsis* sp.

Owing to the cyclic lipopeptide nature of KF (Fig. 1D), this compound could be synthesized either by a nonribosomal peptide synthetase pathway (NRPS) or by a ribosomally synthesized and posttranslationally modified peptide pathway (RiPP) (11, 12). In both cases, it is relatively straightforward to computationally link a biosynthetic gene cluster to its product. In NRPSs, the number of modules, the substrate specificity of the adenylation (A) domains, and the presence or absence of epimerization (E) domains are distinctive for a particular molecule. In RiPPs, the primary amino acid sequence of the final molecule is directly encoded on a precursor peptide within the biosynthetic gene cluster. To examine whether bacterial or fungal members of the *Bryopsis* sp. microbiome encode biosynthetic gene clusters consistent with the structure of KF, we deeply sequenced the *Bryopsis*-2015 metagenomic DNA using Illumina [48 million paired-end reads, 175 base pairs (bps)], performed multiple iterations of assemblies on the produced data, and queried the reads and resulting assemblies for possible biosynthetic gene clusters using several search strategies (see materials and methods).

First, for RiPP pathways, we searched for the expected primary amino acid sequence of KF (VTVVPRITIVFTV, where R replaces ornithine and T replaces dehydrobutyrine) against a database of algal metagenomic reads and contigs using the tBLASTn mode of the Basic Local Alignment Search Tool. Second, for NRPS pathways, we constructed a database of 36 phylogenetically diverse NRPS proteins and used it as a query for tBLASTn searches against the *Bryopsis*-2015 metagenomic assembly (18,308 scaffolds with length >5 kbps). Third, we subjected the same scaffolds to antiSMASH analysis, a stand-alone tool for the unbiased identification of small-molecule

¹Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA. ²Princeton Center for Theoretical Science, Princeton University, Princeton, NJ 08544, USA. ³Center for the Physics of Biological Function, Princeton University, Princeton, NJ 08544, USA. ⁴Institute of Marine and Environmental Technology, University of Maryland Center for Environmental Science, Baltimore, MD 21202, USA. *These authors contributed equally to this work. †Present address: National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Silver Spring, MD 20910, USA. ‡Corresponding author. Email: donia@princeton.edu

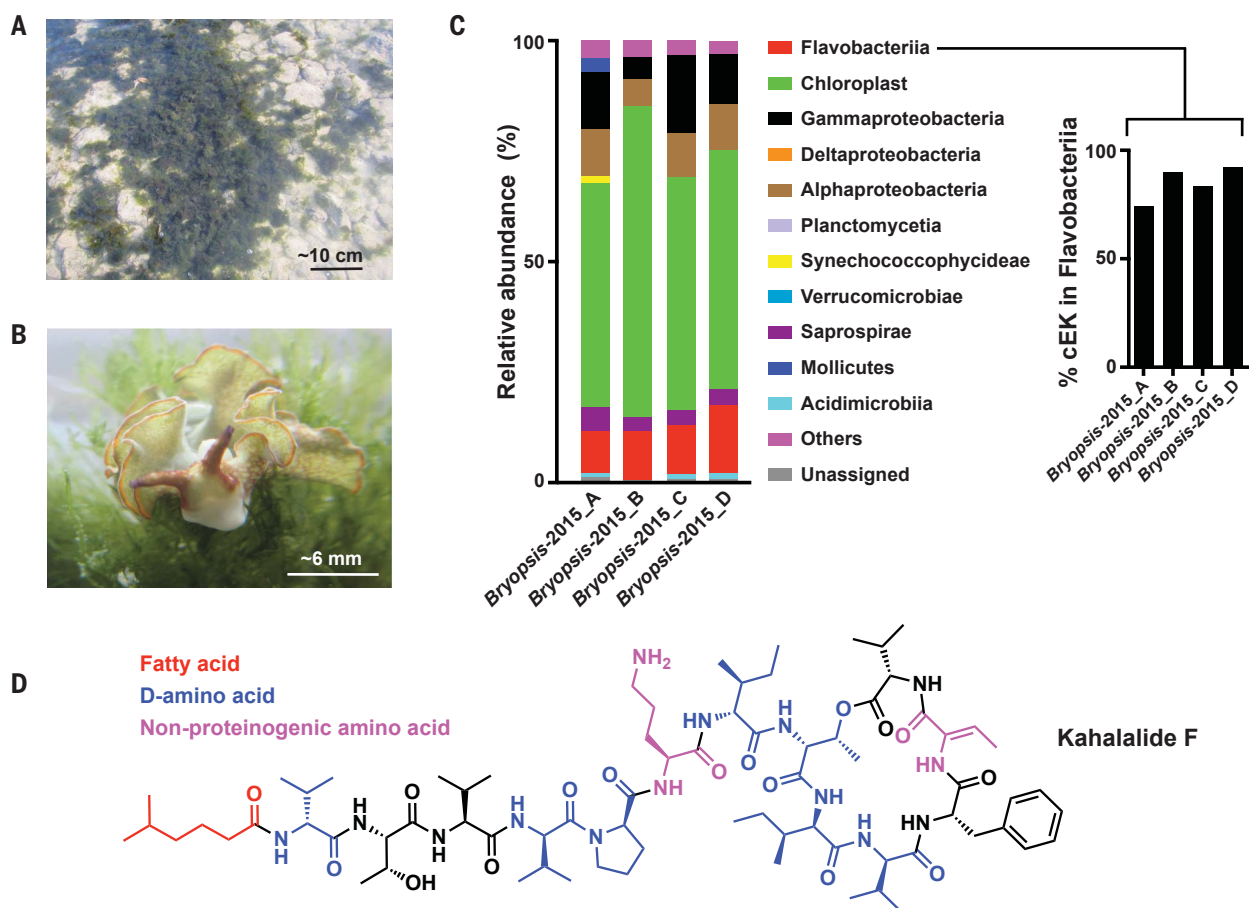


Fig. 1. Bacterial composition and chemical defense of the *Bryopsis*–*Elysia* symbiotic system. (A) The natural *Bryopsis* sp. bloom studied in this work. (B) *E. rufescens* feeding on *Bryopsis* sp. in the laboratory. (C) Composition of *Bryopsis*-2015-associated bacterial communities across four different replicates collected on two consecutive days (A and B: 29 March 2015; C and D: 30 March 2015) and grouped at the class

level. Rare taxa (<1% of total sequences) are labeled as “Others.” The inset bar graph represents the percentage of a single 16S rDNA sequence (hereafter designated “cEK”) that dominates the class Flavobacteriia.

(D) Molecular structure of kahalalide F (KF), the main defensive chemical in both *Bryopsis* sp. and *E. rufescens*. Structural features commonly associated with microbial biosynthesis are depicted in red, blue, and pink.

biosynthetic gene clusters (13, 14). Although no matches to the primary KF sequence were retrieved using the RiPP search strategy, 130 scaffolds (5 to 394 kbps in length) that encode NRPS pathways were discovered using the NRPS search strategy. In addition, antiSMASH annotated a total of 25 scaffolds as NRPSs, only three of which were not detected with the BLAST strategy. Unfortunately, none of the 133 scaffolds containing possible NRPSs constituted a 13-domain pathway, as would be expected for KF, implying that the presumed KF biosynthetic gene cluster may be fragmented or unclustered. Careful analysis of the multimodular-NRPS-containing scaffolds revealed 33 scaffolds that have similar average GC content (~50%) and similar coverage in the metagenome (>300X), suggesting that they may be part of one large NRPS pathway or at least part of the same bacterial genome (fig. S2).

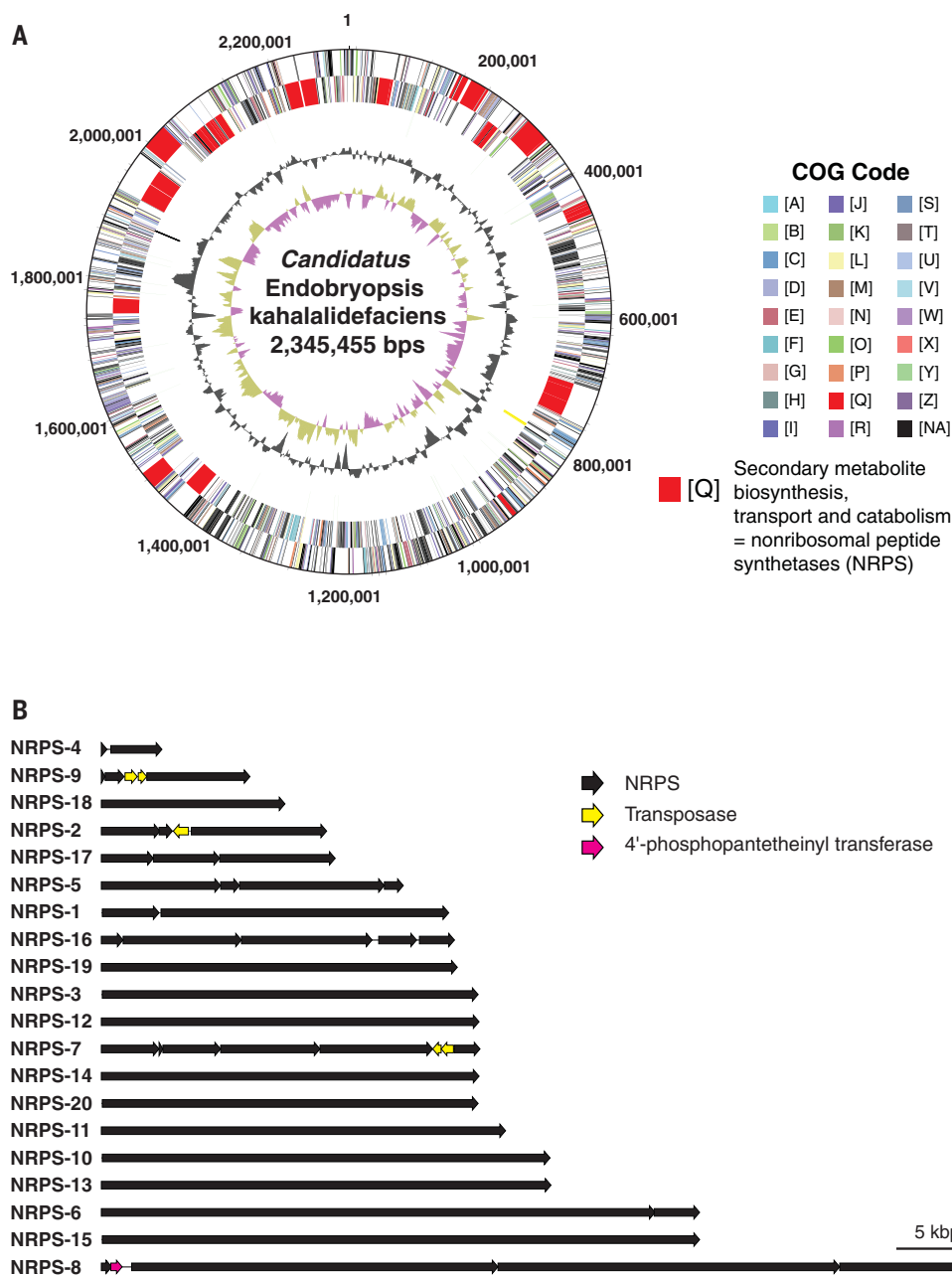
To obtain a better assembly for the fragmented NRPS scaffolds, we subjected *Bryopsis*-2015 meta-

genomic DNA to long-read, single-molecule real-time (SMRT) sequencing using the Pacific Biosciences platform (three SMRT cells, 325,000 reads, median insert size of 9 kbps). After assembly and error correction with the high-quality Illumina reads, we successfully closed a ~2.3-Mbp bacterial genome in which several of the partial NRPS sequences from the Illumina assembly were joined to produce a 55.6-kbp NRPS pathway (NRPS-8) that is consistent with the final KF structure (Fig. 2). Several lines of evidence link NRPS-8 to KF biosynthesis: (i) It begins with a starting condensation domain, as predicted by multiple biosynthetic analysis algorithms (NaPDos and antiSMASH); these types of domains catalyze an amide bond formation between a fatty acid and the first amino acid of the growing peptide chain (13–15). Indeed, the KF structure starts with 5-methyl hexanoic acid conjugated to the first valine residue. (ii) It consists of 13 NRPS modules, as expected for the tridecapeptide KF. (iii) Modules 1, 4, 5, 7, 8, 9, 10, and 12 of NRPS-8

contain epimerization domains, which are responsible for converting the stereochemistry of the loaded amino acid from the L form to the D form (17). In agreement with this, KF harbors 7 D-amino acids at positions 1, 4, 5, 7, 8, 9, and 10, which correspond exactly to these epimerizing modules (threonine residue at position 12 is dehydrated to dehydrobutyrine, which is achiral). (iv) The predicted substrate specificity of almost all adenylation domains in NRPS-8 matches the amino acid residues observed in KF, in an analysis performed using published algorithms and further supported by a detailed phylogenetic analysis (see materials and methods and figs. S3 and S4) (16). Notably, at 40 kbps downstream of NRPS-8, the same genome encodes a 16S rRNA gene that is identical to the aforementioned, most abundant bacterium in the sample. A 16S rDNA-based phylogenetic tree of related organisms placed this bacterium as a previously unknown genus in the family Flavobacteriaceae, order Flavobacteriales, class Flavobacteriia, and

Fig. 2. The large biosynthetic capacity of “*Ca. E. kahalalidefaciens*.”

(A) Circular map of the “*Ca. E. kahalalidefaciens*” chromosome. Tracks (from outermost to innermost) represent genes on the forward strand, genes on the reverse strand, RNAs, GC content, and GC skew. Genes are color coded according to the Cluster of Orthologous Groups (COG) categories in the Integrated Microbial Genome platform (21) (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, posttranslational 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, defense mechanisms; W, extracellular structures; X, mobilome, prophages, and transposons; Y, nuclear structure; Z, cytoskeleton; and NA, not assigned). Note that 99.7% of the aggregate length of genes classified in the COG category Q in the “*Ca. E. kahalalidefaciens*” chromosome (red) correspond to NRPS pathways and occupy 20% of the genome coding capacity. **(B)** Genetic organization of the 20 NRPS pathways in “*Ca. E. kahalalidefaciens*” (ordered by size), where each arrow indicates a single gene.



phylum Bacteroidetes (fig. S5)—a taxonomical group not known as a prolific producer of complex small molecules. We termed this bacterium “*Candidatus Endobryopsis kahalalidefaciens*” (Latin for “kahalalide-making”).

Symbiont biosynthetic capacity

Further analysis of the “*Ca. E. kahalalidefaciens*” genome revealed that it harbors 20 NRPS pathways, ranging in size from 3.9 (NRPS-4) to 55.6 (NRPS-8) kbps and accounting for all 33 high-coverage, ~50% GC NRPS fragments previously found in the Illumina assembly (Fig. 2 and fig. S2). Unusually for bacteria, these NRPS pathways occupy 20% of the genome. Since the initial dis-

covery of KF from *Bryopsis* sp. and *E. rufescens*, other work has detected >15 cyclic and linear lipopeptides of different lengths and amino acid compositions from samples collected at the same site; these molecules are known collectively as the kahalalides (Fig. 3) (8, 9, 17–20). Because these molecules have been reported from the same species (sometimes at abundances similar to that of KF) and because of the structural similarities between KF and the rest of the kahalalides, we asked whether the remaining 19 NRPSs in the “*Ca. E. kahalalidefaciens*” genome encode other known kahalalides.

To answer this question, we compared the 19 remaining NRPSs encoded in the genome

to the kahalalide structures. Overall, we were able to link eight additional NRPSs in the “*Ca. E. kahalalidefaciens*” genome to eight corresponding kahalalide chemotypes (a distinct chemotype can encompass several kahalalides sharing the same amino acid sequence and differing only in the length or hydroxylation of the fatty acid, hydroxylation of a proline residue in the molecule, or linearization of the cyclic peptide) (Fig. 3 and fig. S6). Although a genetic knockout (loss of function) or heterologous expression (gain of function) is typically needed to unequivocally connect biosynthetic gene clusters to their products, five lines of evidence indicate that “*Ca. E. kahalalidefaciens*” synthesizes

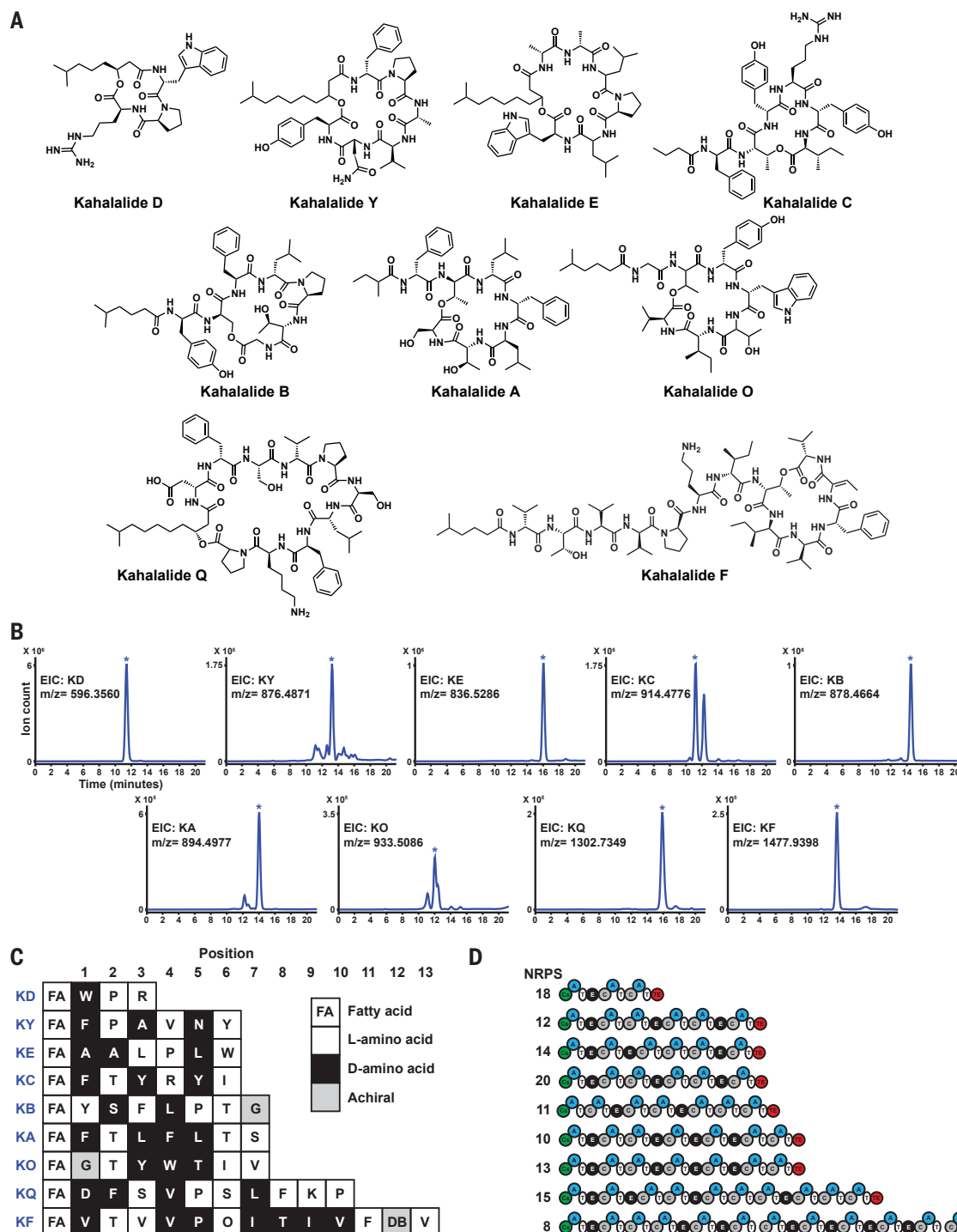


Fig. 3. Structural and biosynthetic diversity of kahalalides from *Bryopsis* sp. (A) Molecular structures of nine kahalalides that can be bioinformatically linked to specific NRPS pathways in “Ca. E. kahalalidefaciens” and chemically detected in *Bryopsis*-2015. (B) Extracted ion chromatograms (EICs) for the mass/charge ratios (m/z) corresponding to the (M + H)⁺ ions of the molecules in (A), as detected in the chemical extract of *Bryopsis*-2015. An asterisk indicates the peak corresponding to the molecule of interest. (C) Amino acid composition of the nine kahalalides shown in (A), indicating the position and stereochemistry of each amino acid on a linear scale. Kahalalide names are represented with two-letter abbreviations (e.g., kahalalide D = KD), and amino acids are denoted by their

canonical one-letter codes (A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr) except for ornithine (O) and dehydrobutyrine (DB). FA, fatty acid. (D) Domain organization of “Ca. E. kahalalidefaciens”-encoded NRPS pathways that were bioinformatically linked to corresponding kahalalides in C (Cs, starting condensation domain; A, adenylation; T, thiolation; E, epimerization; C, condensation; and TE, thioesterase). The number of modules in a given NRPS and the number and position of epimerization domains encoded in it (black) match exactly what is observed in its linked product (see also substrate specificity analyses in figs. S3 and S6 and detailed HR-MS/MS analyses in table S1 and figs. S1 and S7).

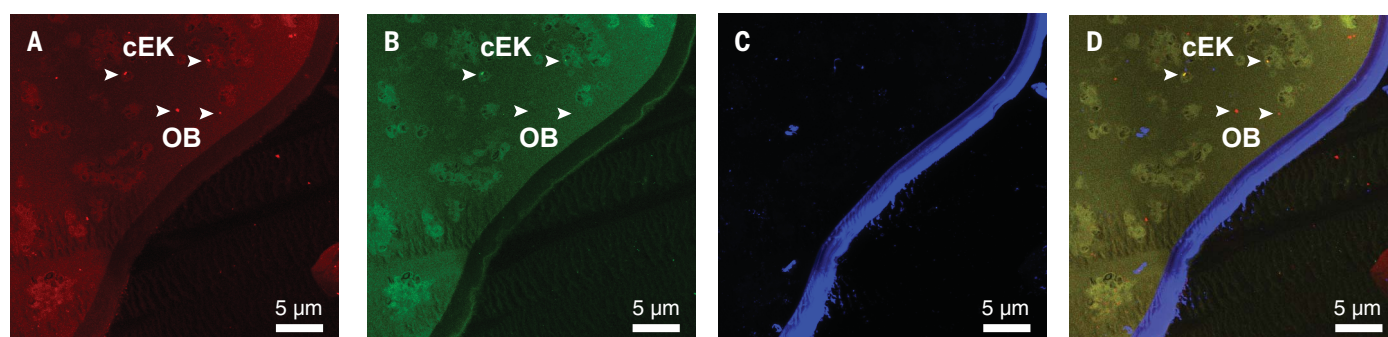


Fig. 4. Localization of “Ca. *E. kahalalidefaciens*” in *Bryopsis* sp. using FISH. Epifluorescent micrographs of a *Bryopsis* sp. section hybridized with (A) universal eubacterial probes EUB338 I to III (labeled with Cy3, red) and (B) “Ca. *E. kahalalidefaciens*”-specific probe JZP2 (labeled with 6-FAM, green). (C) Algal cell wall of the same section viewed under a 4',6'-diamidino-2-phenylindole (DAPI) channel using calcofluor counterstaining. (D) Composite of images (A) to (C), showing the colocalization of the red and green signals for “Ca. *E. kahalalidefaciens*” but not for other bacteria. Arrows indicate bacteria: cEK, “Ca. *E. kahalalidefaciens*”; OB, other bacteria.

all other kahalalides previously reported from this *Bryopsis* sp. First, consistent with the lipopeptide nature of the kahalalides, the matched NRPSs begin with a starting condensation domain. Second, the number of modules in the NRPSs and the position of epimerization domains in the modules match exactly the size and L- and D-amino acid positions in the corresponding molecules. Third, the predicted substrate specificity of adenylation domains in the NRPS (based on prediction algorithms and detailed phylogenetic analysis of the A domains) predominantly match the amino acid composition of the corresponding kahalalides (Fig. 3 and figs. S4 and S6). Fourth, no other NRPS pathways recovered from the metagenome are consistent with the structure of any of the kahalalides (97 scaffolds). Finally, and most importantly, in every case where we were able to bioinformatically match an NRPS in “Ca. *E. kahalalidefaciens*” to a corresponding kahalalide, we successfully confirmed the presence of the exact molecule in the chemical extract of *Bryopsis*-2015 using HPLC–HR-MS/MS analysis (Fig. 3, fig. S7, and table S1). These results establish “Ca. *E. kahalalidefaciens*” as a symbiotic microbial factory for at least nine complex molecules that are abundant enough to be isolated from the algal host. Of these nine molecules, at least one was shown to be essential for the host's chemical defense against predators.

An intracellular symbiont with a reduced genome

Apart from the NRPS-coding regions, the remaining coding capacity of the “Ca. *E. kahalalidefaciens*” genome is only 1.87 Mbps (average genome size in the Flavobacteriaceae family is ~3.7 Mbps), indicating that this bacterium is undergoing evolutionary genome reduction. Next, we annotated the “Ca. *E. kahalalidefaciens*” genome using the Integrated Microbial Genome platform (21) and compared it to the closest free-living and genome-sequenced bacterium: *Mangrovimonas* sp. ST2L12 (88% DNA sequence identity for the 16S rRNA gene) (22) (fig. S5). The genome of “Ca. *E. kahalalidefaciens*” encodes 873 protein-coding genes that can be functionally assigned

into a Cluster of Orthologous Groups (COG) category and 449 that can be classified into a Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, whereas the *Mangrovimonas* sp. ST2L12 genome encodes 1868 and 801 such genes, respectively. Most enzymes and pathways for DNA replication and protein synthesis, cofactor biosynthesis, and central and intermediary metabolism (glycolysis, tricarboxylic acid cycle, pentose phosphate pathway, electron transport, and fatty acid synthesis and degradation) are present in both “Ca. *E. kahalalidefaciens*” and *Mangrovimonas* sp. ST2L12. However, some DNA repair genes and pathways (damaged nucleotide removal and base excision repair, nucleotidyltransferase involved in DNA repair, and RNA helicase), as well as genes involved in chemotaxis, detoxification, and adaptations to atypical conditions, are mostly absent from “Ca. *E. kahalalidefaciens*,” a finding that is commonly observed in obligate symbionts undergoing genome reduction (23) (table S2).

Most notably, no complete pathway to any of the 20 proteinogenic amino acids can be identified in the “Ca. *E. kahalalidefaciens*” genome, with most pathways missing all or the majority of the genes. The only partially complete series are in the early steps of aromatic amino acid biosynthesis leading to chorismate, but the downstream enzymes for phenylalanine, tyrosine, and tryptophan biosynthesis are mostly absent (table S2). Additionally, the “Ca. *E. kahalalidefaciens*” genome shows two features that are rarely found in reduced genomes. First, its GC content has not been reduced: The average GC content of “Ca. *E. kahalalidefaciens*” is 52%, higher than that of *Mangrovimonas* sp. ST2-L12 (38%). Second, it contains an unusually high number (84) of transposons, transposases, and transposable elements. With this enrichment of mobile elements, we wondered whether overt signs of horizontal gene transfer (HGT) can be detected in the genome. We searched for signs of genome heterogeneity by multiple parametric methods, including GC content (24), *k*-mer frequencies (*k* = 2 or 3) (25), and dinucleotide bias (26). We detected a total of seven genomic regions that strongly deviate in their composition from the genome average (fig. S8 and table S3). Six of these re-

gions range in size from 0.1 to 2.6 kbps and encode hypothetical proteins. A seventh region of ~26 kbps (coordinates: 1821316 to 1847296) encodes several phage proteins, including capsid, portal, and tail sheath proteins, and likely constitutes a complete prophage. No other clear signs of HGT were detected across the entire genome, including the 20 NRPSs (fig. S8).

Next, we tested whether “Ca. *E. kahalalidefaciens*” is capable of free living by attempting its cultivation in various media and employing a sensitive, high-throughput sequencing-based method for its detection (see materials and methods and fig. S9). These efforts, however, were not successful, suggesting that “Ca. *E. kahalalidefaciens*” is an obligate symbiont. Indeed, most genome-reduced, obligate symbionts live intracellularly in their hosts (23). To determine whether “Ca. *E. kahalalidefaciens*” is an intracellular or extracellular symbiont of *Bryopsis* sp., we performed fluorescence in situ hybridization (FISH) on fixed *Bryopsis*-2015 specimens using fluorescent probes that target regions of the 16S rRNA. Two sets of probes were used: a combination of universal eubacterial probes EUB338 I, II, and III (27, 28) and one that is specific for “Ca. *E. kahalalidefaciens*” (JZP2; see materials and methods). Both general and specific probes localized “Ca. *E. kahalalidefaciens*” to the cellular compartment of *Bryopsis*, where most detected signals were from dispersed cells within the algal cytosol (Fig. 4). Notably, we found that a partial 16S rRNA gene sequence (with 100% DNA sequence identity to that of “Ca. *E. kahalalidefaciens*”) was previously cloned from a Brazilian sample of *Bryopsis* sp. (fig. S5) (29). This previous study suggested that the bacterium harboring this sequence is also intracellular and widespread in *Bryopsis* sp. living in warm-temperate and tropical environments (30). Finally, we searched a recently published dataset composed of 16S rRNA gene amplicon sequences from 1194 sponges, 37 sediments, and 195 seawater samples for sequences that match “Ca. *E. kahalalidefaciens*” (31). At a cutoff of 97% DNA sequence identity, no matches were found, further supporting the specific, obligate, and intracellular symbiotic relationship between “Ca. *E. kahalalidefaciens*” and *Bryopsis* sp.

Given that there is no clear nutritional advantage provided by “*Ca. E. kahalalidefaciens*” and that 20% of the bacterial genome is dedicated to the biosynthesis of complex small molecules, we assume that the most important aspect of this relationship is chemical defense in exchange for a tolerant intracellular environment in which to live. Notably, “*Ca. E. kahalalidefaciens*” is defective in synthesizing all 16 amino acids that are incorporated as substrates during the biosynthesis of the kahalalides (Fig. 3). These substrates are provided by the autotrophic *Bryopsis* sp. host, although other intracellular bacteria may also contribute. These findings establish an unusual collaborative model for the biosynthesis of the kahalalides, where the host provides the initial substrates and the symbiont provides the biochemical machinery to transform these simple substrates into complex, biologically active molecules that may otherwise be unattainable in algal biochemistry.

A tripartite symbiosis

Through kleptoplasty, *E. rufescens* sequesters and maintains intact and functioning chloroplasts from the cells of its *Bryopsis* sp. diet. Can the mollusk likewise “steal” intact intracellular bacterial symbionts of *Bryopsis* sp. (e.g., “*Ca. E. kahalalidefaciens*”)? To answer this question, we collected several specimens of *E. rufescens*, confirmed that they contain KF, and then used deep 16S rRNA gene amplicon sequencing from whole-animal metagenomic DNA of three individuals (an average of 57,000 reads per sample) to search for “*Ca. E. kahalalidefaciens*.”

The *E. rufescens* microbiome is dominated by sequences matching the algal chloroplast (47.0%, on average), Mollicutes (29.0%), and Gamma-proteobacteria (12.5%) (fig. S10), in agreement with a previous study (32). Despite the sequencing depth performed in our study, we barely detected 16S rRNA gene sequences that match “*Ca. E. kahalalidefaciens*” in *E. rufescens* (an average of 0.06% of the obtained reads). Additional sequencing of 16S rRNA gene amplicons from sections of two coarsely dissected *E. rufescens* individuals (an average of 81,000 reads per section) confirmed that there was no specific symbiosis organ in the mollusk and that “*Ca. E. kahalalidefaciens*” cells appear to be fully digested upon ingestion (fig. S10). Finally, though we could clearly observe signals for other bacteria using general eubacterial probes, negative FISH experiments on several *E. rufescens* sections using a specific probe for “*Ca. E. kahalalidefaciens*” corroborated our sequencing results (fig. S11). Taken together, our results establish chemical but not symbiont sequestration as a likely means of KF acquisition, although the mechanistic details of this step are unknown.

A model for the evolution of defense chemicals

Next, we asked what molecular mechanisms generate the observed diversity in kahalalide structures. Because we found no evidence of HGT-mediated acquisition of NRPS pathways

in the “*Ca. E. kahalalidefaciens*” genome (fig. S8), we looked for signs of increased genome plasticity (33, 34) and intragenome genetic exchange. Briefly, we segmented the genome into 150-bp fragments, produced a whole-genome identity dot plot by comparing all fragments, and searched for genomic loci that contain a high density of matched fragment pairs (>60% sequence identity). On average, NRPS loci showed a 50-fold higher density of matched fragments than the rest of the genome (Fig. 5A; see also materials and methods and fig. S12), in a largely indiscriminate (matched fragments within a single NRPS pathway are neither more frequent nor more similar than those between pathways) (fig. S13) and pervasive (matched fragments cover >98.8% of the combined NRPS sequences) manner. A notable outlier in this analysis was NRPS-8, where matched fragments between NRPS-8 and other pathways showed lower density and percent identity than between any other pairs of NRPSs (Fig. 5A and figs. S12 and S13).

Consecutive matched fragments between pairs of NRPSs form visually distinct stripes on identity dot plots, which represent alignments that start and end with abrupt changes in DNA sequence identity as a result of possible genetic exchange events (Fig. 5B and fig. S12). We identified more than 100 such events: Three represent full pathways and can therefore be explained by typical duplication and divergence events (one between NRPS-10 and NRPS-13; two between NRPS-3, NRPS-12, and NRPS-20), whereas the remaining ones represent only parts of the pathways and are best explained by separate recombination events (Fig. 5C and fig. S14). In most cases, lineage trees for different aligned sequences that appear in the same set of NRPSs do not follow the same path (Fig. 5D), indicating that frequent shuffling substantially disrupts the trajectories of pathway duplication. Finally, 9 of the 20 NRPSs contain or occur within a 5-kbp window of an annotated transposase or transposon (fig. S15), suggesting that transposition is at least partially responsible for the observed plasticity, as seen in other systems (33, 34). These findings motivated us to propose a new model for the diversifying evolution of NRPSs in the “*Ca. E. kahalalidefaciens*” genome, where new pathways arise through duplication and divergence events that are concurrently accompanied by a high frequency of interpathway recombination, resulting in the observed diversity and an intertwined evolutionary history.

Why would a bacterial symbiont produce so much chemical diversity? In other words, are these molecules equally important in establishing a sustainable symbiotic relationship, especially given that maintaining and expressing their large biosynthetic pathways is a major metabolic cost? We reasoned that nonessential pathways should decay and eventually get lost, whereas essential ones should be intact and maintained. We looked for signs of genomic decay in the 20 NRPS pathways. Indeed, NRPS-4 and NRPS-16 show truncated domains and modules, and

NRPS-2, NRPS-7, and NRPS-9 contain domains that have been interrupted by transposases (Fig. 2B). The remaining NRPS pathways appear to be pristine. We then compared the expression levels of the NRPS pathways under native conditions, reasoning that essential pathways would be highly expressed, whereas decaying ones would be expressed at a lower level or not expressed at all. Metatranscriptomic sequencing data from *Bryopsis*-2015 (40 million paired-end reads, 150 bps; 41 million single-end reads, 100 bps) were mapped to the “*Ca. E. kahalalidefaciens*” genome (see materials and methods), resulting in the successful alignment of 2.2 million reads (mostly to possible mRNAs) (Fig. 6A). We then ranked all genes according to their expression level (data S1), which produced a trimodal distribution: About 50% of the genes are almost silent or have a very low level of expression, 45% are at median level, and 5% are among the most highly expressed (Fig. 6B). Comparison of the expression level of the NRPS genes showed a similar distribution. Genes from NRPS-1, -2, -4, -6, -9, -16, and -17 (NRPS-2, -4, -9, and -16 were predicted to be decaying pathways) had the lowest expression levels, and genes from NRPS-8 had the highest. NRPS-8 genes were among the most highly expressed genes in the entire genome, along with cell division and transcription genes. NRPS-8 alone accounted for more than 12% of the transcriptional activity of “*Ca. E. kahalalidefaciens*,” and in total, all of the NRPS pathways accounted for 26% (Fig. 6A). These results establish NRPS-8 (encoding the most toxic kahalalide, KF) as a potentially essential pathway, and NRPS-2, -9, and -16 as potentially on their way to being lost.

We thus propose a model for NRPS pathway generation, evolution, and selection in “*Ca. E. kahalalidefaciens*” (supplementary text). In our model, new pathways are created from an ancestral one through duplication and divergence, followed by extensive and continuous interpathway genetic exchange to generate more diversity. Newly generated pathways are then selected for or against, depending on unknown selection pressures, leading to their maintenance or loss, followed by a continuous repetition of the same cycle (fig. S16).

Discussion

KF is the most toxic of all known kahalalides and is likely the main contributor to *Bryopsis* sp. and *E. rufescens* chemical defense against predators (7). KF and more than 15 related kahalalides have been reported from this predator-prey system, yet the original source of this library of molecules has been unknown. In this study, we used metagenomics, metatranscriptomics, chemical analysis, microscopy, and evolutionary genomics to uncover the molecular details of kahalalide production by an elusive third partner, an intracellular bacterial symbiont of *Bryopsis* sp.

Chemically defended mollusks are widespread in the marine environment, but the source of their defense chemicals varies (5). Although chemical sequestration from diet has been proposed

in many cases (35), de novo biosynthesis has also been demonstrated through isotope labeling experiments (36). Here, we provide evidence that defense chemicals in *E. rufescens* do not

originate in the algal diet itself but in intracellular bacterial symbionts within the alga. A similar indirect acquisition mechanism may be true for other chemicals found in mollusks, such

as those acquired from dietary sponges and tunicates. Most work on symbioses in marine algae has focused on extracellular symbionts, as exemplified by the symbiosis between *Emiliania*

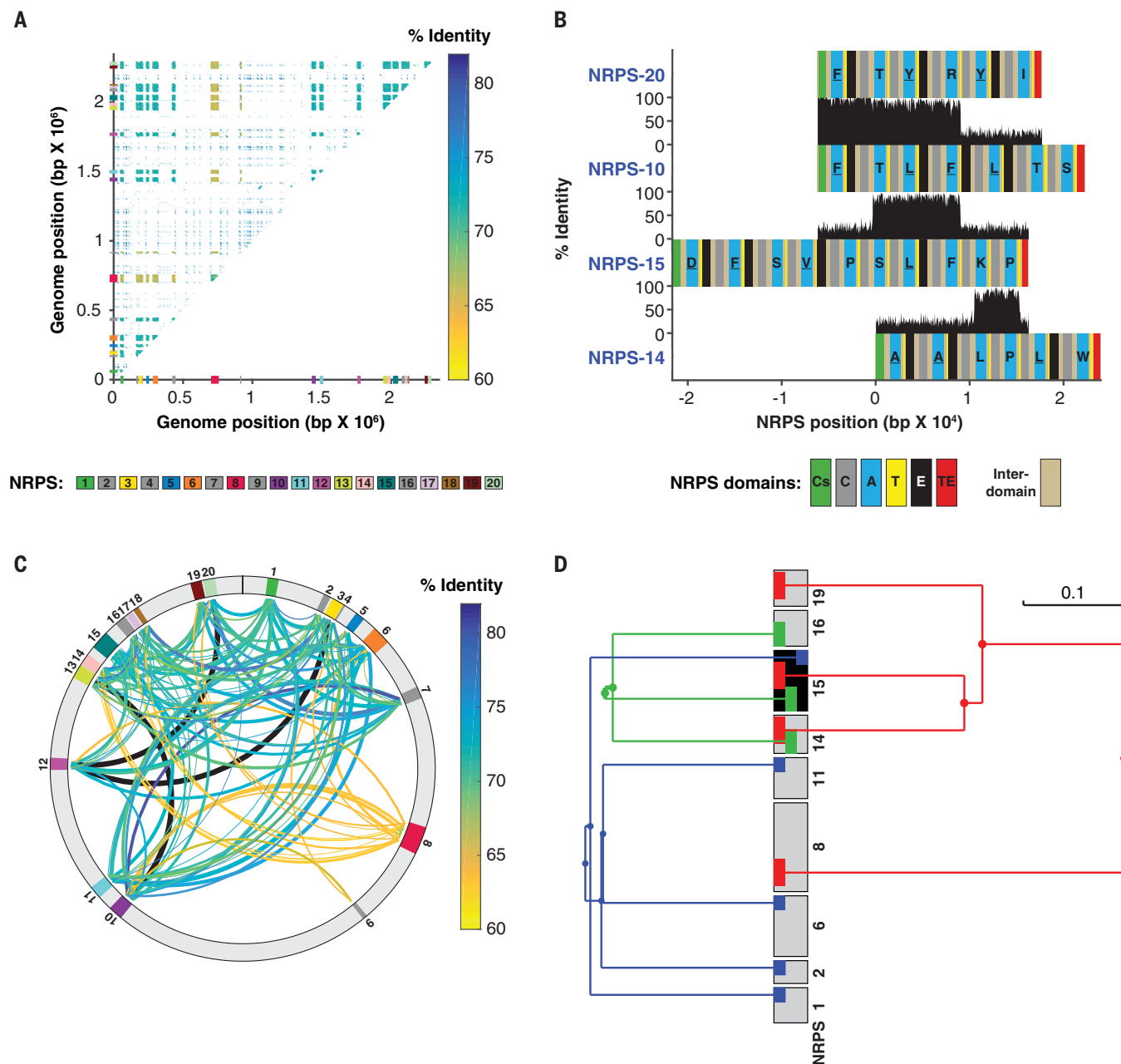


Fig. 5. Intensive genetic exchange in “Ca. *E. kahalalidefaciens*” NRPS pathways. (A) Pairs of sequence fragments sharing more than 60% identity across the whole “Ca. *E. kahalalidefaciens*” genome, shown on a linear scale on both the x and y axes. Diagonal matches were removed, and color indicates percent identity. Positions of the 20 NRPS pathways are indicated on the axes, matching the color code below. (B) Examples of aligned sequences between NRPS pathways. Note the abrupt rise and fall in sequence identity. Colors in pathways indicate different domains, matching the color code below. Encoded amino acids are indicated by their single-letter abbreviations, and underlined letters indicate D-amino acids. (C) Circular representation of “Ca. *E. kahalalidefaciens*” genome recombination events (coordinate one of the genome is indicated by a solid black line), in which NRPS pathways are shown at their respective positions around the genome and connecting colored

lines indicate identified pairwise recombination events. Lines follow a yellow-to-blue color code that represents the percent identity between pairs of aligned sequences, and their thickness positively correlates with the length of the sequence. Sequences shared with NRPS-8 and NRPS-9 have a lower percent identity than ones shared with other pathways. Black lines indicate the three identified duplication events. (D) Phylogenetic trees for three groups of aligned sequences (red, green, and blue) that together cover the entire NRPS-15. Although the three aligned sequences are found in NRPS-15, they do not follow the same phylogenetic lineage and are likely a result of three independent recombination events. NRPS-15 is shown as a black rectangle, and NRPS pathways that share one or more of the three groups of aligned sequences with it are shown as gray rectangles. The scale bar at top right indicates 10% sequence divergence.

huxleyi and *Phaeobacter inhibens*, in which *P. inhibens* influences the biology of the host through alternating symbiotic and pathogenic cycles (37), and that between *Ulva* sp. and several bacteria, in which the symbionts trigger proper host growth and development (38). Here, we describe the functional importance of intracellular symbionts of marine macroalgae for the ecology of their host, which has been previously overlooked. We also establish macroalgae as an additional group of marine organisms in which bacterial symbionts produce complex natural products discovered in the host, as has been previously observed in marine tunicates, sponges, bryozoans, and crustaceans (4).

The diversifying evolution of NRPS pathways in the “*Ca. E. kahalalidefaciens*” genome is particularly noteworthy because it reveals a strategy for chemical innovation in bacteria in a HGT-independent manner. The closest examples are two cases reported in bacterial symbionts of marine ascidians: one in which the same polyketide synthase pathway has duplicated multiple times in the same genome, presumably resulting in higher expression levels (39), and another in which a cyanobactin pathway has duplicated once and accumulated strategic point mutations in its precursor peptide to produce novel molecules (40, 41). The number of biosynthetic pathways (20 NRPSs) and resulting molecules discovered in “*Ca. E. kahalalidefaciens*” is substantial, especially given its reduced genome and strict intracellular lifestyle, and approaches the diversity encoded in the genome of *Entotheonella* sp., an extracellular symbiont of marine sponges and a talented producer of natural products (42, 43).

We do not yet understand the nature of the selective pressures on the “*Ca. E. kahalalidefaciens*” genome. Assuming that all kahalalides have roles in chemical defense, we propose either that they contribute to a synergistic cocktail, where molecules function additively on diverse predators or pathogens and the selection acts to maintain diversity, or that each molecule has a specific role and is selected for independently. There is some support for the latter hypothesis in that KF has been shown to have a strong antipredatory activity (7), and the rest of the kahalalides have low to no cytotoxicity or antimicrobial activity and their natural role has not yet been determined (9). Furthermore, levels of different kahalalides produced in the same system vary considerably between times of collection (17), suggesting that they are selected for individually and not globally. We also do not understand how *E. rufescens* sequesters, concentrates, and resists the cytotoxicity of the kahalalides. This is complicated by the fact that their molecular target is still unknown, even for KF, which has been evaluated as a promising anticancer agent in several human clinical trials (44). We show that chemical and not symbiont sequestration is responsible for kahalalide acquisition in *E. rufescens*, but exactly how and where this occurs await further investigation. These outstanding questions are not specific to *E. rufescens* and the kahalalides but are true for almost all marine mollusks that sequester toxins from their diet.

Materials and methods summary

An expanded description of materials and methods can be found in the supplementary materials.

Bryopsis sp. and *E. rufescens* collection and processing

Bryopsis sp. and *E. rufescens* samples were collected from Black Point Bay, Honolulu, Hawaii (N21°15'34"; W157°47'24"). Small portions of *Bryopsis* sp. and individual *E. rufescens* specimens were preserved frozen for chemical analysis, preserved in RNAlater (ThermoFisher Scientific, USA) for DNA and RNA analysis, or fixed in paraformaldehyde for FISH analysis.

Metagenomic DNA and RNA extraction and sequencing

Total metagenomic DNA used for both Illumina and Pac Bio sequencing was extracted from lyophilized *Bryopsis* sp. and *E. rufescens* using the Mo Bio Power Biofilm DNA isolation kit (Mo Bio Laboratories, USA; now Qiagen, USA). Total RNA was extracted from RNAlater-preserved *Bryopsis* sp. using MasterPure Complete DNA and RNA Purification Kit (Epicentre, USA; now Illumina, USA). Illumina sequencing was performed on an Illumina HiSeq 2500 sequencer (Illumina, USA), and Pac Bio sequencing was performed on a PacBio RS II instrument (Pacific Biosciences, USA).

Fluorescence in situ hybridization (FISH)

Fixed specimens of *Bryopsis* sp. were embedded in LR White resin (Electron Microscopy Sciences, USA), cut into 2-μm-thick slices, hybridized with either universal eubacterial probes or a “*Ca. E. kahalalidefaciens*”-specific probe, counterstained using Calcofluor White Stain (Sigma Aldrich, USA), and imaged on a Leica SP8 confocal microscope.

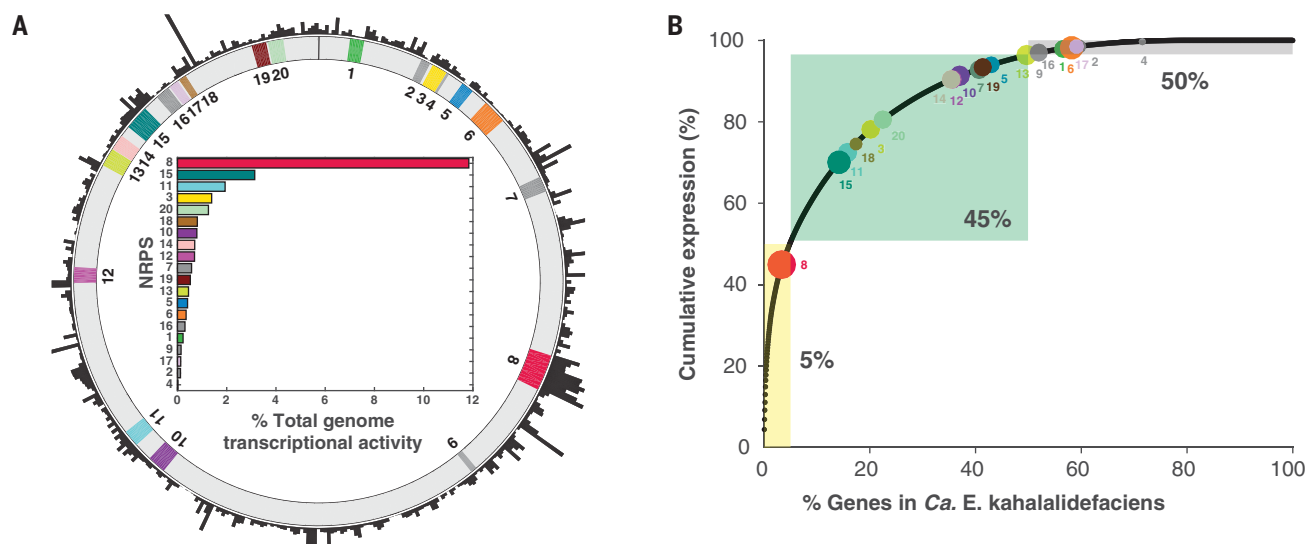


Fig. 6. Metatranscriptomic analysis of “*Ca. E. kahalalidefaciens*.” (A) Genome-wide transcriptional activity of “*Ca. E. kahalalidefaciens*” using metatranscriptomic data collected from *Bryopsis*-2015. Black bars indicate total counts for each mapped position and averaged per 5 kbps of the genome. The inset bar graph indicates the ranked contribution of each NRPS pathway as a percentage of the total genome transcriptional

activity. (B) “*Ca. E. kahalalidefaciens*” genes (x axis, shown in percentage of the total number of genes in the genome) plotted against their cumulative expression (y axis, shown in percentage of the total expression of all genes in the genome). Note the trimodal distribution of genes according to their expression and that genes from the NRPS pathways appear in all three modes.

Bacterial cultivation and screening for “*Ca. E. kahalalidefaciens*”

Freshly collected *Bryopsis* sp. (5 g) was homogenized in 9 ml of sterile IX Artificial Sea Water (Instant Ocean Sea Salts, 35 g/liter, Instant Ocean, USA) using a mortar and pestle, and the algal homogenate was serially diluted in a 10-fold series. 100-μl aliquots of each dilution were plated on eight different media in triplicate, and plates were incubated at room temperature for 1 to 2 weeks. Colonies were scraped from one plate of each of the eight media, and DNA was extracted from the mixture and screened for the presence of “*Ca. E. kahalalidefaciens*” using high-throughput 16S rRNA gene amplicon sequencing.

Chemical extraction and analysis of *Bryopsis*-2015

Lyophilized *Bryopsis* sp. was extracted with methanol and analyzed by HPLC–HR-MS/MS using an Agilent 6500 Series Q-TOF LC/MS system (Agilent, USA).

Evolutionary analysis of the “*Ca. E. kahalalide*” genome and the 20 NRPSs

The “*Ca. E. kahalalidefaciens*” genome was segmented into fragments of 150 bps in length, with a 30-bp sliding window. Each fragment was aligned to all other fragments to obtain their percent identity; pairs of fragments with identities higher than 60% were recorded as “matched.” Consecutive matched fragments between two genomic regions form stripes with a slope of 1 and varying intercepts on identity dot plots. These stripes were then used to infer genetic exchange events throughout the genome.

REFERENCES AND NOTES

1. V. J. Paul, K. E. Arthur, R. Ritson-Williams, C. Ross, K. Sharp, Chemical defenses: From compounds to communities. *Biol. Bull.* **213**, 226–251 (2007). doi: [10.2307/25066642](#); pmid: [18083964](#)
2. J. Piel, Metabolites from symbiotic bacteria. *Nat. Prod. Rep.* **21**, 519–538 (2004). doi: [10.1039/b310175b](#); pmid: [15282634](#)
3. J. Piel, Metabolites from symbiotic bacteria. *Nat. Prod. Rep.* **26**, 338–362 (2009). doi: [10.1039/b703499g](#); pmid: [19240945](#)
4. M. Morita, E. W. Schmidt, Parallel lives of symbionts and hosts: Chemical mutualism in marine animals. *Nat. Prod. Rep.* **35**, 357–378 (2018). doi: [10.1039/C7NP00053G](#); pmid: [29441375](#)
5. M. T. Davies-Coleman, in *Molluscs: From Chemo-ecological Study to Biotechnological Application*, G. Cimino, M. Gavagnin, Eds. (Springer, 2006), pp. 133–157.
6. M. T. Hamann, P. J. Scheuer, A Bioactive Dipeptide from the Sacoglossan Mollusk *Elysia rufescens* and the Green Alga *Bryopsis* sp. *J. Am. Chem. Soc.* **115**, 5825–5826 (1993). doi: [10.1021/ja00066a061](#)
7. M. A. Becerro, G. Goetz, V. J. Paul, P. J. Scheuer, Chemical defenses of the sacoglossan mollusk *Elysia rufescens* and its host Alga *bryopsis* sp. *J. Chem. Ecol.* **27**, 2287–2299 (2001). doi: [10.1023/A:1012287105923](#); pmid: [11817082](#)
8. M. T. Hamann, C. S. Otto, P. J. Scheuer, D. C. Dunbar, Kahalalides: Bioactive Peptides from a Marine Mollusk *Elysia rufescens* and Its Algal Diet *Bryopsis* sp. *J. Org. Chem.* **61**, 6594–6600 (1996). doi: [10.1021/jo960877+](#); pmid: [11667527](#)
9. J. Gao, M. T. Hamann, Chemistry and biology of kahalalides. *Chem. Rev.* **111**, 3208–3235 (2011). doi: [10.1021/cr100187n](#); pmid: [21480581](#)
10. K. N. Pelletreau et al., Sea slug kleptoplasty and plastid maintenance in a metazoan. *Plant Physiol.* **155**, 1561–1565 (2011). doi: [10.1104/pp.111.174078](#); pmid: [21346171](#)
11. M. A. Fischbach, C. T. Walsh, Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic, machinery, and mechanisms. *Chem. Rev.* **106**, 3468–3496 (2006). doi: [10.1021/cr0503097](#); pmid: [16895337](#)
12. P. G. Arison et al., Ribosomally synthesized and post-translationally modified peptide natural products: Overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **30**, 108–160 (2013). doi: [10.1039/C2NP20085F](#); pmid: [23165928](#)
13. T. Weber et al., antiSMASH 3.0: a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **43**, W237–W243 (2015). doi: [10.1093/nar/gkv437](#); pmid: [25948579](#)
14. K. Blin et al., antiSMASH 4.0: improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **45**, W36–W41 (2017). doi: [10.1093/nar/gkx319](#); pmid: [28460038](#)
15. N. Ziemert et al., The natural product domain seeker NaPDoS: A phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLOS ONE* **7**, e34064 (2012). doi: [10.1371/journal.pone.0034064](#); pmid: [22479523](#)
16. D. Baranasić et al., Predicting substrate specificity of adenylation domains of nonribosomal peptide synthetases and other protein properties by latent semantic indexing. *J. Ind. Microbiol. Biotechnol.* **41**, 461–467 (2014). doi: [10.1007/s10295-013-1322-2](#); pmid: [24104398](#)
17. K. V. Rao et al., Kahalalides V-Y isolated from a Hawaiian collection of the sacoglossan mollusk *Elysia rufescens*. *J. Nat. Prod.* **71**, 772–778 (2008). doi: [10.1021/np070508g](#); pmid: [18407693](#)
18. Y. Kan, T. Fujita, B. Sakamoto, Y. Hokama, H. Nagai, Kahalalide K: A new cyclic depsipeptide from the hawaiian green alga *Bryopsis* species. *J. Nat. Prod.* **62**, 1169–1172 (1999). doi: [10.1021/np990053y](#); pmid: [10479330](#)
19. G. Goetz, Y. Nakao, P. J. Scheuer, Two Acyclic Kahalalides from the Sacoglossan Mollusk *Elysia rufescens*. *J. Nat. Prod.* **60**, 562–567 (1997). doi: [10.1021/np970045m](#)
20. F. D. Horgen et al., A new depsipeptide from the sacoglossan mollusk *Elysia ornata* and the green alga *Bryopsis* species. *J. Nat. Prod.* **63**, 152–154 (2000). doi: [10.1021/np990402o](#); pmid: [10650101](#)
21. A. Checucci, A. Mengoni, The integrated microbial genome resource of analysis. *Methods Mol. Biol.* **1231**, 289–295 (2015). doi: [10.1007/978-1-4939-1720-4_18](#); pmid: [25343872](#)
22. B. Dinesh et al., Comparative genome analyses of novel *Mangrovimonas*-like strains isolated from estuarine mangrove sediments reveal xylan and arabinan utilization genes. *Mar. Genomics* **25**, 115–121 (2016). doi: [10.1016/j.margen.2015.12.006](#); pmid: [26795059](#)
23. J. P. McCutcheon, N. A. Moran, Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* **10**, 13–26 (2011). doi: [10.1038/nrmicro2670](#); pmid: [22064560](#)
24. S. Garcia-Vallvé, A. Romeu, J. Palau, Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome Res.* **10**, 1719–1725 (2000). doi: [10.1101/gr.130000](#); pmid: [11076857](#)
25. F. Zhou, V. Olman, Y. Xu, Barcodes for genomes and applications. *BMC Bioinformatics* **9**, 546 (2008). doi: [10.1186/1471-2105-9-546](#); pmid: [19091119](#)
26. S. Karlin, Global dinucleotide signatures and analysis of genomic heterogeneity. *Curr. Opin. Microbiol.* **1**, 598–610 (1998). doi: [10.1016/S1369-5274\(98\)80095-7](#); pmid: [10066522](#)
27. R. I. Amann et al., Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl. Environ. Microbiol.* **56**, 1919–1925 (1990). pmid: [2200342](#)
28. H. Daims, A. Brühl, R. Amann, K. H. Schleifer, M. Wagner, 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). doi: [10.1016/S0723-2020\(99\)80053-8](#); pmid: [10553296](#)
29. J. Hollants et al., Who is in there? Exploration of endophytic bacteria within the siphonous green seaweed *Bryopsis* (Bryopsidales, Chlorophyta). *PLOS ONE* **6**, e26458 (2011). doi: [10.1371/journal.pone.0026458](#); pmid: [22028882](#)
30. J. Hollants, F. Leliaert, H. Verbruggen, O. De Clerck, A. Willems, Host specificity and coevolution of Flavobacteriaceae endosymbionts within the siphonous green seaweed *Bryopsis*. *Mol. Phylogenet. Evol.* **67**, 608–614 (2013). doi: [10.1016/j.ympev.2013.02.025](#); pmid: [23499613](#)
31. T. Thomas et al., Diversity, structure and convergent evolution of the global sponge microbiome. *Nat. Commun.* **7**, 11870 (2016). doi: [10.1038/ncomms11870](#); pmid: [27306690](#)
32. J. Davis et al., Characterization of the bacterial community of the chemically defended Hawaiian sacoglossan *Elysia rufescens*. *Appl. Environ. Microbiol.* **79**, 7073–7081 (2013). doi: [10.1128/AEM.01568-13](#); pmid: [24014539](#)
33. T. A. Hendry et al., Ongoing Transposon-Mediated Genome Reduction in the Luminous Bacterial Symbionts of Deep-Sea Ceratioid Anglerfishes. *mBio* **9**, e01033-18 (2018). doi: [10.1128/mBio.01033-18](#); pmid: [29946051](#)
34. W. Qi et al., *Ca. Endozoicomonas cretensis*: A Novel Fish Pathogen Characterized by Genome Plasticity. *Genome Biol. Evol.* **10**, 1363–1374 (2018). doi: [10.1093/gbe/evy092](#); pmid: [29726925](#)
35. J. R. Pawlik, M. R. Kernan, T. F. Molinski, M. K. Harper, D. J. Faulkner, Defensive chemicals of the Spanish dancer nudibranch *Hexabranchus sanguineus* and its egg ribbons: Macrolides derived from a sponge diet. *J. Exp. Mar. Biol. Ecol.* **119**, 99–109 (1988). doi: [10.1016/0022-0981\(88\)90225-0](#)
36. J. Kubanek, R. J. Andersen, Evidence for de novo biosynthesis of the polyketide fragment of dialuolsterol A by the northeastern pacific doris nudibranch *dialuula sandiegensis*. *J. Nat. Prod.* **62**, 777–779 (1999). doi: [10.1021/np9804839](#); pmid: [10346969](#)
37. A. R. Bramucci et al., The Bacterial Symbiont *Phaeobacter inhibens* Shapes the Life History of Its Algal Host *Emiliania huxleyi*. *Front. Mar. Sci.* **5**, 188 (2018). doi: [10.3389/fmars.2018.00188](#)
38. F. Ghaderiadekani, J. C. Coates, T. Wichard, Bacteria-induced morphogenesis of *Ulva intestinalis* and *Ulva mutabilis* (Chlorophyta): A contribution to the lottery theory. *FEMS Microbiol. Ecol.* **93**, fix094 (2017). doi: [10.1093/femsec/fix094](#); pmid: [28810708](#)
39. J. Lopera, I. J. Miller, K. L. McPhail, J. C. Kwan, Increased Biosynthetic Gene Dosage in a Genome-Reduced Defensive Bacterial Symbiont. *mSystems* **2**, e00096-17 (2017). doi: [10.1128/mSystems.00096-17](#); pmid: [29181447](#)
40. M. S. Donia et al., Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. *Nat. Chem. Biol.* **2**, 729–735 (2006). doi: [10.1038/nchembio829](#); pmid: [17086177](#)
41. M. S. Donia, W. F. Fricke, J. Ravel, E. W. Schmidt, Variation in tropical reef symbiont metagenomes defined by secondary metabolism. *PLOS ONE* **6**, e17897 (2011). doi: [10.1371/journal.pone.0017897](#); pmid: [21445351](#)
42. M. C. Wilson et al., An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* **506**, 58–62 (2014). doi: [10.1038/nature12959](#); pmid: [24476823](#)
43. T. Mori et al., Single-bacterial genomics validates rich and varied specialized metabolism of uncultivated *Entotheonella* sponge symbionts. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 1718–1723 (2018). doi: [10.1073/pnas.1715496115](#); pmid: [29439203](#)
44. B. Miguel-Lillo, B. Valenzuela, J. E. Peris-Ribera, A. Soto-Matos, J. J. Pérez-Ruix, Population pharmacokinetics of kahalalide F in advanced cancer patients. *Cancer Chemother. Pharmacol.* **76**, 365–374 (2015). doi: [10.1007/s00280-015-2800-1](#); pmid: [26093949](#)

ACKNOWLEDGMENTS

We are indebted to J. Davidson and the late Ruth Gates at HIMB for laboratory space and assistance during the field work conducted in this study. We thank G. Laevsky, the Molecular Biology Confocal Microscopy Facility (a Nikon Center of Excellence), J. Yan, and B. Bassler for assistance with confocal microscopy; W. Wang and the Lewis Sigler Institute sequencing core facility for assistance with high-throughput Illumina and 16S rRNA gene amplicon sequencing; L. Tallon, L. Sadzewicz, and the Institute for Genome Sciences, Genomics Resource Center, University of Maryland, for Pac Bio sequencing; M. Cahn and J. Lopez for assistance with metagenomic data analysis; M. T. Hamann, W. L. Cheung-Lee, and A. J. Link for valuable scientific insights about the project; S. Chatterjee for general assistance; and the rest of the Donia lab for useful discussions. **Funding:** Funding for this project has been provided by Princeton University. M.S.D. is funded by an NIH Director's New Innovator Award (ID 1DP2AI124441), and Z.L. is supported by Princeton Center for Theoretical Science and Center for the Physics of Biological Function, the National Science Foundation Physics Frontier Center grant through the Center for the Physics of Biological Function (PHY-1734030), and an NSF

grant (PHY-1607612). **Author contributions:** M.S.D., J.Z., Z.L., and R.T.H. designed the study. J.Z., Z.L., M.D.T., J.D., and M.S.D. performed experiments and analyzed the data. M.S.D., J.Z., Z.L., and M.D.T. wrote the manuscript. J.D. and R.T.H. edited the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. The “*Ca. E. kahalalidefaciens*” genome has been deposited to the

Integrated Microbial Genomes (Joint Genome Institute, U.S. Department of Energy) public repository under IMG submission ID 115642.

SUPPLEMENTARY MATERIALS
science.sciencemag.org/content/364/6445/eaaw6732/suppl/DC1
Materials and Methods

Supplementary Text
Figs. S1 to S16
Tables S1 to S3
References (45–58)
Data S1

15 January 2019; accepted 7 May 2019
10.1126/science.aaw6732

A microbial factory for defensive kahalalides in a tripartite marine symbiosis

Jindong Zan, Zhiyuan Li, Ma. Diarey Tianero, Jeanette Davis, Russell T. Hill and Mohamed S. Donia

Science **364** (6445), eaaw6732.
DOI: 10.1126/science.aaw6732

A little help from a friend

The Hawaiian sea slug *Elysia rufescens* grazes on an alga called *Bryopsis* sp. The alga defends itself from predators using peptide toxins decorated with fatty acids, called kahalalides. Zan *et al.* wondered if a third party was involved in toxin production (see the Perspective by Mascuch and Kubanek). Within the alga, a species of bacterium with a very reduced genome was discovered to be a factory for the nonribosomal assembly of a family of kahalalides. The authors elucidated the pathways for generating this chemical diversity. It seems that the sea slug not only tolerates the toxins but, to protect itself from being eaten by fish, grazes on the alga to accumulate kahalalide.

Science, this issue p. eaaw6732; see also p. 1034

ARTICLE TOOLS

<http://science.sciencemag.org/content/364/6445/eaaw6732>

SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2019/06/12/364.6445.eaaw6732.DC1>

RELATED CONTENT

<http://science.sciencemag.org/content/sci/362/6418/1034.full>

REFERENCES

This article cites 57 articles, 7 of which you can access for free
<http://science.sciencemag.org/content/364/6445/eaaw6732#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2019 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works