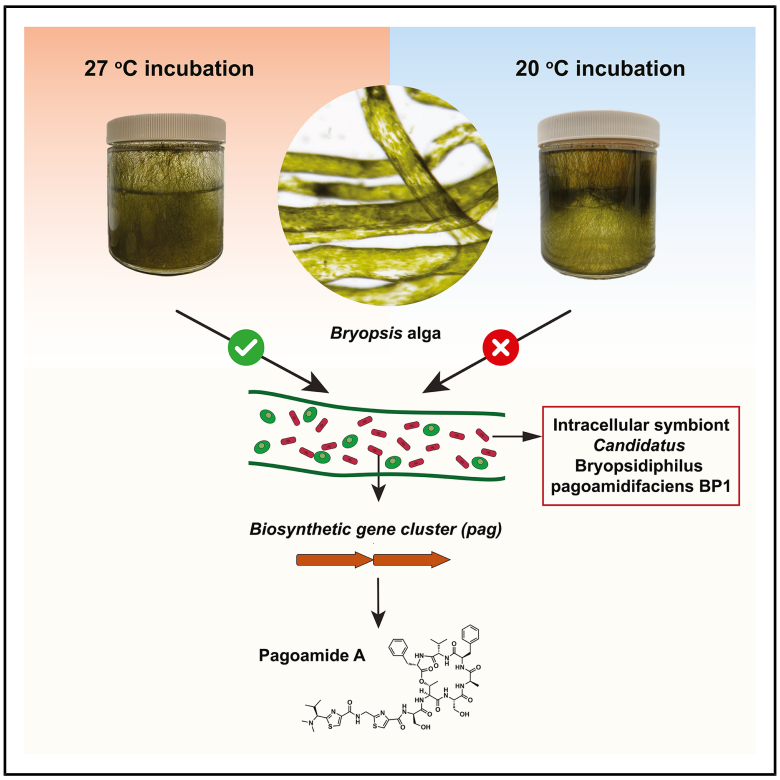
Article

Current Biology

Environmentally controlled production of pagoamide A in marine macroalgae by an intracellular bacterial symbiont

Graphical abstract



Authors

Jie Liu, Eugenia Glukhov, Olivier De Clerck, William H. Gerwick, Mohamed S. Donia

Correspondence

donia@princeton.edu

In brief

Liu et al. show that pagoamide A, a depsipeptide isolated from the marine macroalga *Bryopsis* sp., is produced by its intracellular bacterial symbiont *Candidatus Bryopsidiphilus pagoamidifaciens* in a temperature-dependent manner. *Ca.* B. pagoamidifaciens symbionts are associated with marine macroalgae from diverse geographical locations and show specific genomic variations.

Highlights

- The algal symbiont *Candidatus Bryopsidiphilus pagoamidifaciens* produces pagoamide A
- *Ca.* B. pagoamidifaciens is an intracellular symbiont with a highly reduced genome
- The load of *Ca.* B. pagoamidifaciens in its *Bryopsis* host is temperature dependent
- Genomically diverse *Ca.* B. pagoamidifaciens strains are widespread in marine algae

Liu et al., 2026, Current Biology *36*, 63–79
January 5, 2026
https://doi.org/10.1016/j.cub.2025.11.023

CellPress

Current Biology

CellPress

Article

Environmentally controlled production of pagoamide A in marine macroalgae by an intracellular bacterial symbiont

Jie Liu,[1] Eugenia Glukhov,[2] Olivier De Clerck,[3] William H. Gerwick,[2,4] and Mohamed S. Donia[1,5,6,7,*]
[1]Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA
[2]Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093, USA
[3]Biology Department, Research Group Phycology, Ghent University, 9000 Ghent, Belgium
[4]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093, USA
[5]Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08544, USA
[6]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA
[7]Lead contact
*Correspondence: donia@princeton.edu
https://doi.org/10.1016/j.cub.2025.11.023

SUMMARY

Marine algae are a rich source of diverse molecules, most of which are thought to be produced by the alga itself. We recently reported the discovery of pagoamide A from a cultured marine macroalga collected from American Samoa. Here, we found that the production of pagoamide A is conditional upon environmental temperature. Using comparative metagenomic, metatranscriptomic, and metabolomic analyses of algal cultures, we identified a nonribosomal peptide synthetase biosynthetic gene cluster (NRPS BGC) in the algal microbiome that varies in abundance between producing and non-producing conditions and whose architecture and biosynthetic logic match pagoamide A (named *pag*). *pag* belongs to a bacterium that we named "*Candidatus* Bryopsidiphilus pagoamidifaciens BP1," a new genus in the family Amoebophilaceae and a relative of amoeba, arthropod, and nematode endosymbionts. *Ca.* B. pagoamidifaciens lives intracellularly in its *Bryopsis* sp. algal host, harbors a reduced genome (1.7 Mbp), has lost most genes essential for free living, and is enriched in genes containing eukaryotic domains. By quantitatively monitoring longitudinal algal cultures under varying conditions for 9 weeks, we found that the abundance of both *Ca.* B. pagoamidifaciens and pagoamide A undergoes dramatic fluctuations in response to temperature changes. Finally, we discovered three additional strains of *Ca.* B. pagoamidifaciens that vary in their NRPS BGCs and eukaryotic domain-containing genes from algal samples of diverse geographical origins. Our findings suggest that symbiont-derived production of algal molecules is more common than previously anticipated and provide a unique case of environmental control of both symbiont and chemical levels in marine algae.

INTRODUCTION

Macroalgae play a crucial role in coastal marine ecosystems for their contribution to nutrient cycling, carbon sequestration, and providing a food source or habitat for a plethora of species.[1] Microbial symbionts of marine macroalgae are strongly integrated into the host's biology and ecology by exchanging important nutrients[2,3]; affecting algal morphology, development, and reproduction[4–8]; and contributing to host defense against pathogens and predators.[9,10] For humans, macroalgal holobionts represent valuable sources for bioactive compounds with industrial and pharmaceutical applications, including polysaccharides, phytopigments, polyunsaturated fatty acids, and phenolic compounds, as well as molecules with antimicrobial, anticancer, and anti-inflammatory properties.[11–14] The chemical composition of such holobionts varies widely and is known to depend on both the algal species involved and the surrounding environmental factors, e.g., salinity, water temperature, nutrient content, and light intensity.[14–16] Bacterial isolates from macroalgal holobionts have also been shown to be rich producers of bioactive molecules, yet their contribution to the chemical composition of environmental samples needs further studies.[9,13,17,18]

One example that we recently reported is that of the giant-celled marine macroalga, *Bryopsis* sp., which is chemically defended by a library of lipopeptides called the kahalalides. We showed that the true producer of the kahalalides is an intracellular bacterial symbiont, *Candidatus* Endobryopsis kahalalidifaciens, in a clear example of how a single symbiont can have a major impact on the chemical composition of a macroalgal holobiont.[19] In search of similar cases, we focused on pagoamide A, a thiazole-containing cyclic depsipeptide that we discovered from the laboratory culture of a marine macroalga collected from American Samoa (Figures 1A and 1B).[20] Based on its structural features (e.g., thiazole moieties, macrocyclization, and D-amino acids), we hypothesized that pagoamide A,

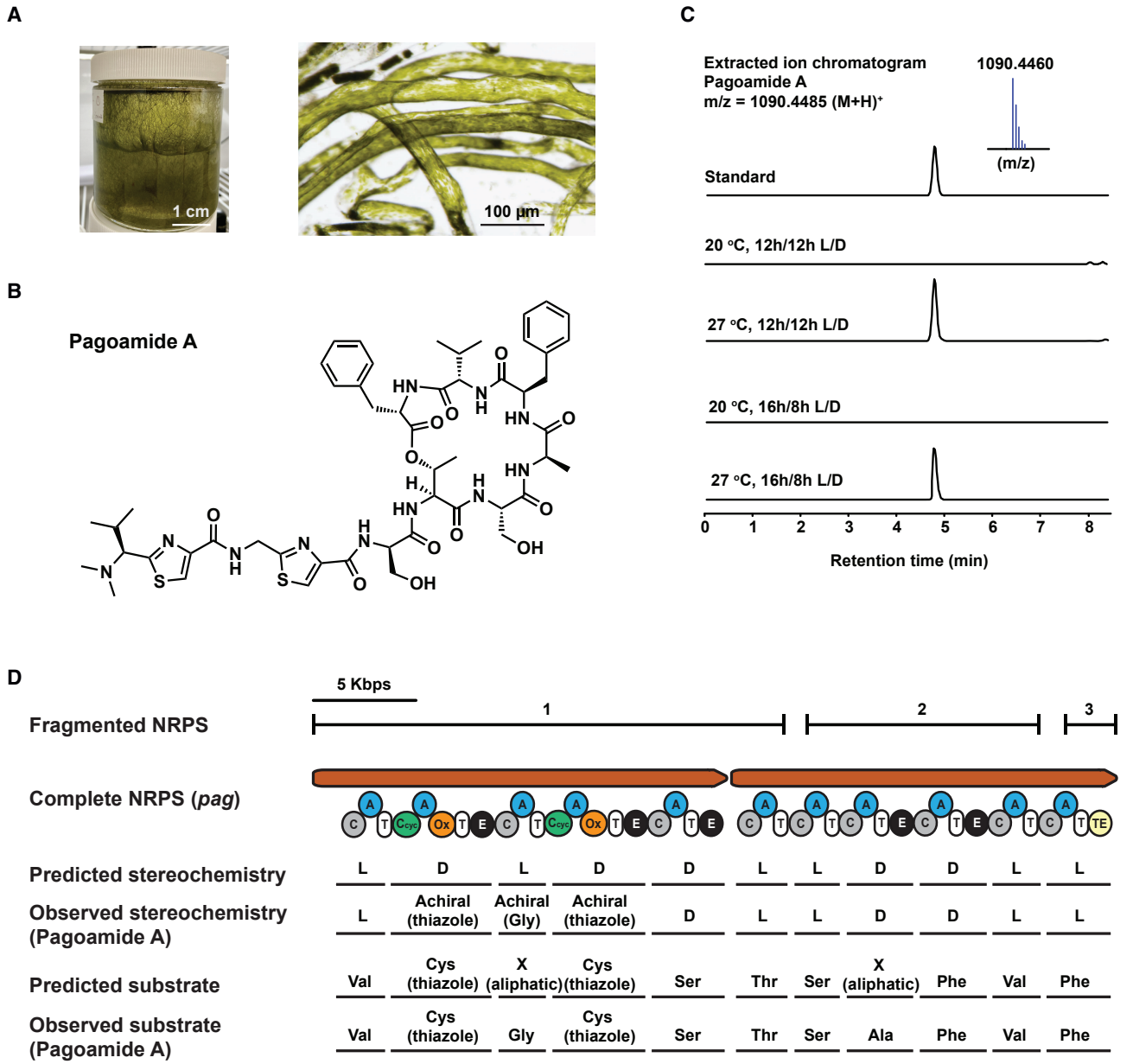


Figure 1. Conditional production of pagoamide A in the holobiont of *Bryopsis* sp. ASF14 algae leads to the identification of *pag*

(A) Left: photograph of *Bryopsis* sp. ASF14 algae cultivated in a glass jar. Right: micrograph of cultivated *Bryopsis* sp. ASF14 algae.

(B) Molecular structure of pagoamide A.

(C) HPLC-HRMS analysis of standard pagoamide A and of chemical extracts obtained from *Bryopsis* sp. ASF14 cultivated under different environmental conditions. An extracted ion chromatogram of pagoamide A, m/z = 1,090.4485 $(M + H)^+$, is shown for each sample.

(D) Genetic and domain architecture of *pag* reveals its concordance with the observed pagoamide A chemical structure. Domain abbreviations: C, condensation; A, adenylation; T, thiolation; Ccyc, heterocyclization; Ox, oxidase; E, epimerization; and TE, thioesterase. On top, initial contigs harboring fragmented *pag* and identified from differential abundance analysis performed on Illumina-sequenced metagenomes and metatranscriptomes are shown with labels 1–3, which are consistent with labels in Figure S1F.

See also Figure S1.

like the kahalalides, is produced by a microbial symbiont rather than the macroalgal host. Bioactivity studies against the H-460 human lung cancer cell line showed that pagoamide A has no cytotoxicity,[20] while antimicrobial assays performed on a panel of bacteria (*Staphylococcus aureus*, *Pseudomonas baumannii*, *Enterococcus faecalis*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Enterobacter* spp., *Bacillus subtilis*, and *Escherichia coli*) and the yeast *Candida albicans* showed that it exerts moderate antibacterial activity against *Bacillus subtilis* only, with a minimum inhibitory concentration of 64 µg/mL (the highest concentration

tested).[21] In light of these results, the exact ecological role of pagoamide A in its natural setting is still unknown.

RESULTS

Pagoamide A production is conditional

Before investigating the biosynthetic origin of pagoamide A, we sought to confirm that cultures of *Bryopsis* sp. ASF14 (corrected from its morphology-based original assignment as *Derbesia* sp. ASF14; see below) continue to produce the molecule after being passaged for many generations in the laboratory (Figure 1A). To that end, we cultured the alga in small scale (50 mL) under two conditions that we regularly use for propagating marine algae (condition 1, 27°C, 16/8 h light/dark diurnal cycle; condition 2, room temperature [20°C ± 2°C], 12/12 h light/dark diurnal cycle) and prepared chemical extracts from each for high-performance liquid chromatography coupled with high-resolution mass spectrometry (HPLC-HRMS) analysis (STAR Methods). Surprisingly, while both conditions supported healthy growth of the algae and produced similar biomass, we could only detect pagoamide A from algae cultured under condition 1 (Figure 1C). To determine whether the change in cultivation temperature or diurnal cycle is the main determining factor for pagoamide A production, we repeated the same experiment but only varied the temperature (27°C or 20°C) while keeping constant the same light/dark cycle (16/8 h light/dark). Once again, we detected pagoamide A only in cultures cultivated at 27°C but not 20°C despite similar algal growths (Figures S1A–S1C), confirming that varying environmental temperature is sufficient to toggle on and off pagoamide A production (Figure 1C). Taken together, these initial results suggest that pagoamide A production in the *Bryopsis* sp. ASF 14 holobiont is conditional and that it is dependent on permissible environmental temperatures.

Conditional production of pagoamide A allows the identification of its biosynthetic origin

Our initial discovery that pagoamide A production is conditional provided us with the ideal set of samples for investigating its biosynthetic origin: virtually a natural pair of "wild-type" and "knockout" cultures. We reasoned that a clear difference in the abundance or expression of the biosynthetic gene cluster (BGC) responsible for producing pagoamide A should be apparent when comparing pagoamide A-positive (PA+) and pagoamide A-negative (PA−) cultures. To test this hypothesis, we deeply sequenced the metagenomes and metatranscriptomes of a pair of PA+ and PA− algal holobiont samples and used them as the bases for our analysis (Illumina, 2X250 bps, 39.7 M and 44.5 M raw paired-end reads for metagenomes, and 2X150 bps, 28.5 M and 40.3 M raw paired-end reads for metatranscriptomes, respectively) (STAR Methods).

Based on the structural features of pagoamide A (Figure 1B), namely the presence of non-proteinogenic amino acids (e.g., D-amino acids, thiazole moieties, and an *N*-methylated residue), as well as the macrocyclization into a depsipeptide, we hypothesized that it is synthesized by either a nonribosomal peptide synthetase (NRPS) biosynthetic pathway or a ribosomally synthesized and posttranslationally modified peptide (RiPP) biosynthetic pathway.[22,23] Thus, we started our analysis by assembling the metagenomic sequencing data from the PA+ culture and searching for BGCs belonging to these two classes that can potentially encode for pagoamide A production. For RiPPs, we directly searched our assembled metagenome for the primary amino acid sequence of pagoamide A (VCGCSTSAFVF) using tBLASTn and found no identical hits. For NRPSs, we used antiSMASH 5,[24] a comprehensive tool for the identification and annotation of BGCs from genomic and metagenomic data. We discovered a total of 109 BGCs from PA+ metagenomic data, twenty-three of which were predicted to contain complete or partial NRPS pathways (Figure S1D). Next, we calculated the abundance of each identified NRPS BGC (measured in reads per kbps per million of sequenced reads, RPKM) in PA+ and PA− samples at both the metagenomic and metatranscriptomic levels and searched for BGCs that are enriched in PA+ cultures (PA+ RPKM/PA− RPKM > 10 or PA− RPKM = 0 at both the metagenomic and metatranscriptomic levels) (Figure S1E). Six partial NPRS BGCs passed this criterion. By using GC content and metagenomic abundance in the PA+ sample as the bases for binning (Figure S1F), we were able to group the six differential BGCs into two distinct clusters, suggesting that they originate from two different genomic sources.

To improve the assembly of these six partial NRPS BGCs, we extracted high molecular weight DNA from the PA+ sample and subjected it to long-read, single-molecule real-time metagenomic sequencing (Pacific Biosciences, 1.9 M HiFi reads, average length 15,682 bps). Indeed, using the long-read-based metagenomic assembly, we found that the six partial NRPS BGCs belonged to two separate circularized bacterial genomes. Notably, we found that the BGC from the higher abundance and lower GC content group (43%–48% GC) is fully consistent with pagoamide A production. This BGC consists of 11 NRPS modules (pagoamide A is composed of 11 amino acids); modules 2 and 4 encode heterocyclization and oxidation domains that catalyze the heterocyclization and oxidation of cysteine, serine or threonine residues into thiazoles or oxazoles (pagoamide A harbors two thiazoles at positions 2 and 4); modules 5, 8, and 9 encode epimerization domains, which are responsible for converting amino acid residues from the L to D configuration (pagoamide A harbors three D-amino acids at positions 5, 8, and 9); and finally, computational prediction of the amino acid substrate specificity of the 11 adenylation domains in the NRPS BGC perfectly matches the pagoamide A structure (Figure 1D). Interestingly, the N terminus of pagoamide A is doubly methylated, yet no methyltransferase gene is found in the identified BGC, suggesting that the responsible enzyme may be encoded elsewhere in the genome or in a different organism altogether (i.e., post-assembly-line modification). With the complete BGC assembled, we repeated the quantification of its abundance in the PA+ and PA− samples. We computed 106-fold enrichment in the PA+ versus PA− sample on the metagenomic level and 80-fold enrichment on the metatranscriptomic level. Taken together, the PA+ enrichment of the identified BGC, coupled with its near-perfect match with the pagoamide A structure, strongly suggests that it is indeed responsible for pagoamide A biosynthesis. Thus, we termed this NRPS BGC *pag*.

Pagoamide A is produced by a genome-reduced, intracellular bacterial symbiont

Next, we focused on the genome harboring *pag*: a complete, circular bacterial genome of 1,698,248 bps in length belonging to the Bacteroidota (formerly Bacteroidetes) phylum (Figure 2A). Using a

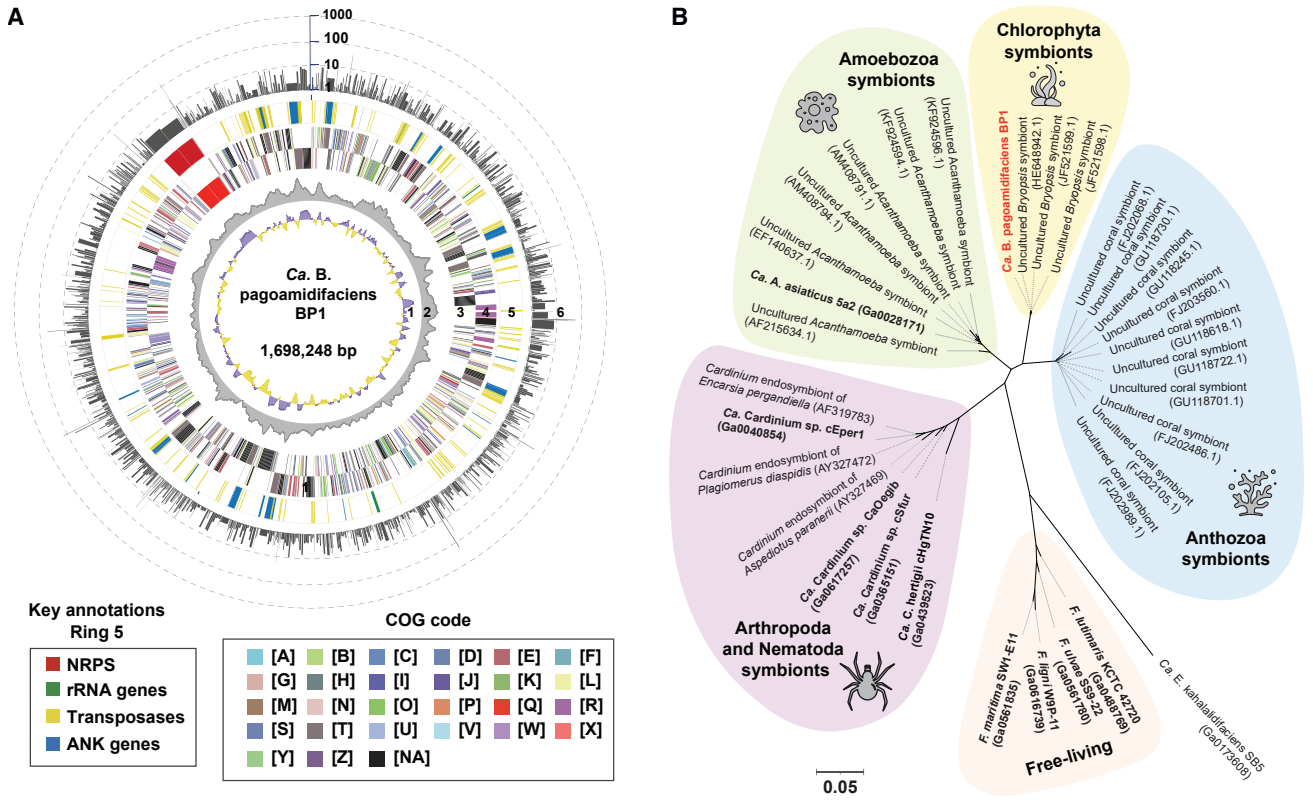


Figure 2. Genome and phylogenetic analysis of *Ca.* B. pagoamidifaciens BP1

(A) Circular genome of *Ca.* B. pagoamidifaciens BP1, retrieved from the metagenome of *Bryopsis* sp. ASF14. Concentric rings (from inside to outside) indicate 1, GC skew; 2, GC content; 3, genes on forward frames; 4, genes on reverse frames; 5, highlighted gene groups of interest; and 6, gene expression level in PA+ sample, representing the RPKM of each gene in exponential scale. The RPKM reference axis indicates 10, 100, and 1,000 from inside to outside. Only genes with RPKM $\geq$ 1 are shown. Genes shown in rings 3 and 4 are classified to their respective Cluster of Orthologous Group (COG) categories based on the Integrated Microbial Genome platform (IMG) annotation (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).

(B) Phylogenetic tree of *Ca.* B. pagoamidifaciens based on the 16S rRNA gene sequence, with selected Amoebophilaceae strains and amplicons from diverse hosts. The phylogenetic tree was constructed using PhyML with 500 bootstraps.[30] Complete genomes used in downstream genome comparison analyses are highlighted in bold (see below).

See also Data S3.

phylogenetic analysis based on the 16S rRNA gene, we classified this bacterium as a new genus in the family Amoebophilaceae, order Cytophagales (Figure 2B). Family Amoebophilaceae only includes two other classified genera: *Candidatus* Amoebophilus spp., which are mainly endosymbionts of *Acanthamoeba* sp.,[25] and *Candidatus* Cardinium, which are endosymbionts of arthropods and nematodes.[26–29] Current members of this family with available complete genomes, i.e., *Ca.* Amoebophilus and *Ca.* Cardinium, are phylogenetically distant (92% 16S rRNA gene identity) from the *pag*-encoding bacterium and do not encode any *pag* homologs.[25–29] On the other hand, the 16S rRNA gene of the *pag*-containing bacterium shows high DNA sequence identity (>99%) to a group of 16S rRNA gene amplicon sequences derived from marine *Bryopsis* algae, suggesting a new clade of algal symbionts within the family Amoebophilaceae (Figure 2B). Finally, the *pag*-containing genome is similarly reduced in size as that of other endosymbionts in the same family—0.9–1.8 Mbps (about one-third of the closest free-living bacterium, *Fulvivirga lutimaris* KCTC 42720, 4.9 Mbp) (Figure 2B)—and lacks many pathways that are essential for free living (see below).

With the genome reduction characteristics of the *pag*-encoding bacterium and its phylogenetic relatedness to other Amoebophilaceae endosymbionts, we hypothesized that it too lives an intracellular symbiotic lifestyle within its algal host. To test this hypothesis, we designed a specific probe targeting the 16S rRNA sequence of the *pag*-encoding bacterium and used it in fluorescence *in situ* hybridization (FISH) experiments performed on intact algal tubular strands from a PA+ culture (STAR Methods). In addition to the specific probe, we also used a universal eubacterial probe to localize all bacterial

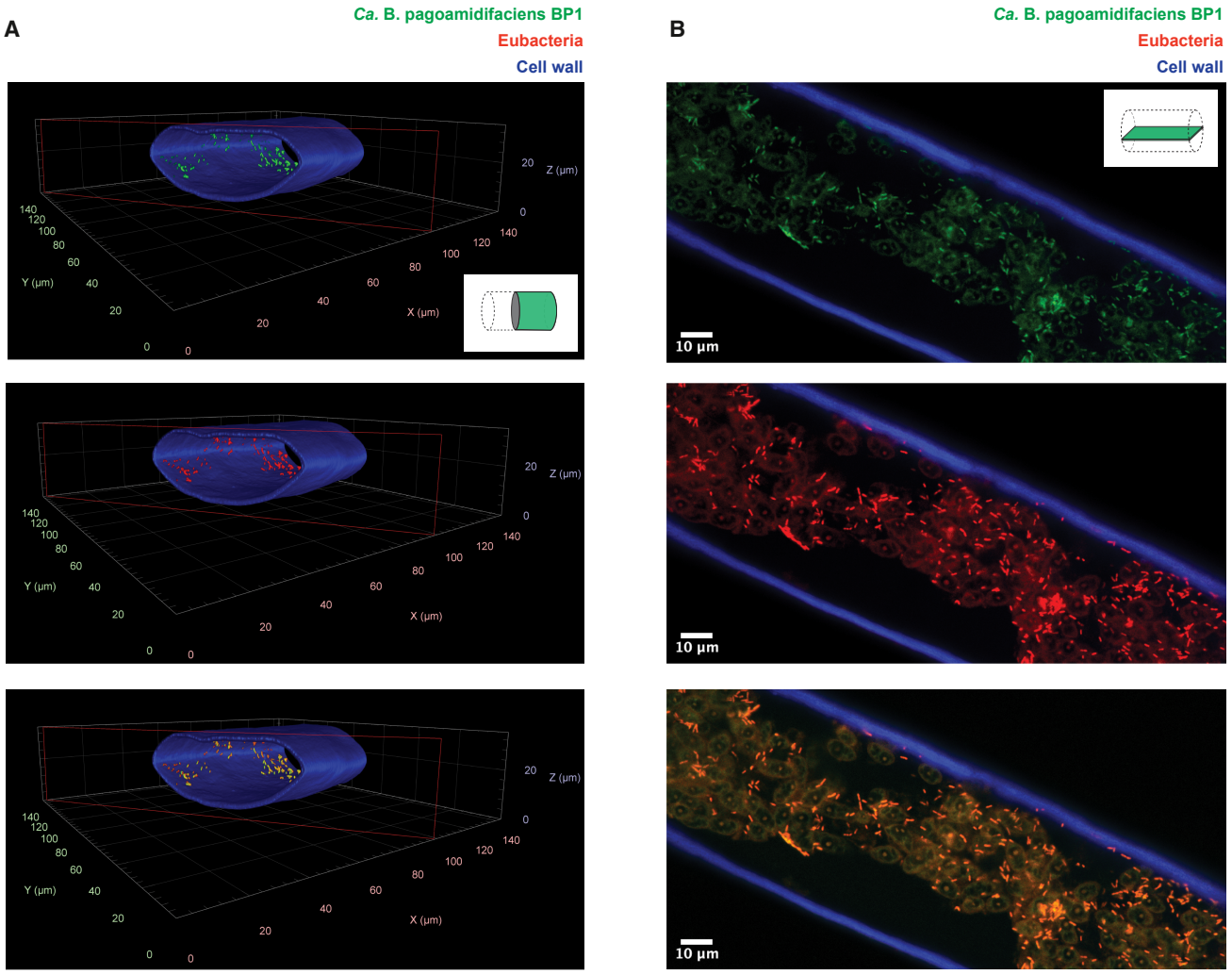


Figure 3. Localization of *Ca.* B. pagoamidifaciens BP1

(A) Three-dimensional volumetric micrographs of a PA+ *Bryopsis* sp. ASF14 algal strand. Orthogonal projections were generated from z stack imaging of the entire strand (depth ~35 μm) with a step size of 0.5 μm.

(B) Representative z-stack micrographs for the same strand shown in A, where each image was generated from a z-stack maximum intensity projection of three consecutive optical sections and a 0.5 μm step.

In both (A) and (B), epifluorescence microscopy was performed after specimens were hybridized with universal eubacterial probes EUB338 I-III (labeled with Cy3, red) and *Ca.* B. pagoamidifaciens BP1 specific probe JL39-892R (labeled with 6-FAM, green) and counterstained with Calcofluor to visualize the algal cell wall (blue) (STAR Methods). The top panel in (A) and (B) shows *Ca.* B. pagoamidifaciens BP1-specific signals detected in the green channel (488 nm), together with the cell wall stain in the blue channel (405 nm). The embedded schematic illustrates the cylindrical algal strand with a cutaway section plane indicating the regions shown in the micrographs. Middle panel shows signals from the universal eubacterial probes detected in the red channel (561 nm), together with the cell wall stain in the blue channel (405 nm). The bottom panel shows a composite of the three channels (green, red, and blue).

See also Figure S2.

members of the algal microbiome.[31] As expected, while the universal probe labeled bacterial cells both inside and on the outer surface of the algal cell (Figure S2), the specific probe clearly localized the *pag*-encoding bacterium to its intracellular compartment (Figure 3). Interestingly, through these localization studies, we found that the *pag*-encoding bacterium can reach high abundance per algal strand (Figures 3 and S2). Taken together, our imaging, taxonomical, and genomic results show that the *pag*-producing bacterium belongs to a new lineage of Amoebophilaceae symbionts that live intracellularly in marine algae. Given the information about pagoamide A production, the intracellular lifestyle of the symbiont, and the taxonomical classification of the host alga into *Bryopsis* sp. (STAR Methods; Figure S3; Data S1), we termed the *pag*-encoding bacterium "*Candidatus* Bryopsidiphilus pagoamidifaciens BP1" (Latin for "*Bryopsis* inhabitant, pagoamide-making," strain BP1).

Ca. B. pagoamidifaciens BP1's genome is highly reduced yet enriched with genes involved in symbiont-host interaction

Next, to shed light on its lifestyle and metabolic capabilities, we annotated the *Ca.* B. pagoamidifaciens BP1's genome using the

Integrated Microbial Genome platform (IMG) (Figure 2A). In comparison with the closest free-living bacterium (*Fulvivirga lutimaris* KCTC 42720, 4,858,936 bps, 4,427 protein-coding genes), the genome of *Ca.* B pagoamidifaciens BP1 (1,698,248 bps, 1,488 protein-coding genes) is extremely reduced and lacks essential free-living traits, including all complete amino acid metabolism pathways, major cofactor metabolism pathways, and central carbohydrate, sulfur, and nitrogen metabolism pathways (Data S2A). It is common that reduced bacterial genomes encode a reduced number of transcription factors.[32] Consistent with this fact, only 1.75% of the *Ca.* B. pagoamidifaciens BP1 genome is annotated as transcription factors, as compared with 3.84% of the genome in *Fulvivirga lutimaris* (Data S2A). The prevalence of transposases in obligate symbionts is known to be associated with genome reduction, as well as genome plasticity and adaptation to the host and intracellular environments.[19,33,34] Indeed, the *Ca.* B. pagoamidifaciens BP1's genome is enriched with transposable elements, which account for 10.84% of the genome coding capacity and 7.11% of total *Ca.* B. pagoamidifaciens BP1 transcripts, compared with 0.03% of the genome in free-living *Fulvivirga lutimaris* (Data S2A). This extreme genome reduction and enrichment in transposable elements strongly suggest that *Ca.* B. pagoamidifaciens BP1 is highly adapted to an obligate intracellular lifestyle and that it is metabolically dependent on its algal host.[35]

To further explore whether Ameobophilaceae symbionts share a common adaptation strategy to different hosts, we conducted a comprehensive comparative genomic analysis between six Amoebophilaceae symbionts (*Ca.* B. pagoamidifaciens BP1, *Ca.* A. *asiaticus* 5a2, and four *Ca.* Cardinium strains) and four free-living *Fulvivirga* strains (STAR Methods; Figure 4A).[36] A total of 382 core orthologous groups (core OGs) are shared between Amoebophilaceae symbionts and *Fulvivirga* strains, representing approximately 27% of the total number of genes in *Ca.* B. pagoamidifaciens BP1. On the other hand, we found that only 28 OGs are consistently shared by all Amoebophilaceae but not all *Fulvivirga* strains (symbiont-core OGs), which are enriched in functions related to amino acid transport and metabolism (cluster of orthologous group [COG] category E) (Data S2A). As these symbionts have lost most genes involved in amino acid metabolism, these symbiont-core genes may play a role in nutrient acquisition. Notably, this set also includes a gene encoding for an ATP:ADP antiporter, which is known to facilitate ATP uptake from eukaryotic hosts into symbiotic bacteria in exchange for ADP, as shown with *Rickettsiae* and *Chlamydiae* spp.[37] In *Ca.* B. pagoamidifaciens BP1, the ATP:ADP antiporter (IMG Gene ID 8013676511) ranks 34th out of all 1,488 genes in its transcriptional level, suggesting the importance of energy scavenging from the algal host for *Ca.* B. pagoamidifaciens BP1 (Data S3).

By contrast, a large proportion of symbiont genes are found only in a subset of symbionts (symbiont-specific OGs) or are unique to one symbiont (symbiont-unique OGs), highlighting the genomic divergence among Amoebophilaceae symbionts. For example, previous studies revealed that the *Ca.* A. *asiaticus* 5a2 genome encodes a type VI secretion system (T6SS).[40,41] We found that T6SS-related genes are present in the four *Ca.* Cardinium strains but not in *Ca.* B. pagoamidifaciens BP1, indicating a lineage variation in secretion systems within the family Amoebophilaceae.

Both *Ca.* Amoebophilus sp. and *Ca.* Cardinium hertigii cHgTN10 genomes are enriched in genes encoding eukaryotic domain-containing proteins.[25,28] One notable example of this category is genes encoding for ankyrin repeat (ANK)-containing proteins, a common structural motif in eukaryotic proteins that is involved in many essential cellular functions by modulating protein-protein interactions.[42] ANK-containing proteins are more prevalent in obligate intracellular bacteria, e.g., *Wolbachia* sp. and *Rickettsia* sp., than facultative host-associated and free-living bacteria,[42,43] and have been directly linked to important microbe-host interactions in symbiotic systems.[44,45] In our analysis, we identified 37 genes encoding ANK-containing proteins in the *Ca.* B. pagoamidifaciens BP1's genome, varying in length from 243 to 12,978 bps and accounting in total for 6.5 % of the genome coding capacity and 2.8% of its transcriptional activity in the PA+ metatranscriptome (Data S3). Strikingly, some of these 37 genes encode an extremely high number of ANK repeats, up to 107 in one case (IMG Gene ID 8013676850) (Figure 4B).

In contrast to this expansion, the genome of *Ca.* Cardinium sp. CaOegib encodes very few eukaryotic domain-containing proteins (Data S2). Similarly, not all analyzed Amoebophilaceae symbiont genomes are depleted in the number of genes encoding for transcription factors as expected for genome-reduced symbionts. For example, 5.68% and 3.58% of the genes in *Ca.* A. *asiaticus* 5a2 and *Ca.* Cardinium sp. cEper1 genomes encode for transcription factors, respectively, which is comparable to the percent found in free-living *Fulvivirga* strains (3.84%). Taken together, the observed variations across Amoebophilaceae symbiont genomes likely reflect their divergent adaptations to distinct host environments.

Ca. B. pagoamidifaciens BP1 is the most transcriptionally active member in the PA+ algal microbiome

After we identified *Ca.* B. pagoamidifaciens BP1 and highlighted the symbiotic features of its genome, we sought to quantify the relative abundance and expression of this organism as a fraction of the *Bryopsis* sp. ASF14 microbiome under the PA+ and PA− conditions. To broadly describe the algal microbiome in our samples, we performed comprehensive binning analysis on metagenomic assemblies generated from long-read and short-read sequencing data to generate metagenome-assembled genomes (MAGs)[46] and complemented this analysis with 16S rRNA gene amplicon sequencing for bacterial community profiling (STAR Methods). In total, we obtained 49 unique and high-quality MAGs, with an average completeness of 97% and an average contamination of 0.6% (29 MAGs are circular, five contain single contigs, and the rest contain multiple contigs) (Figure 5A; Data S4A). The number of MAGs retrieved from metagenomic binning is comparable to the results of our 16S rRNA amplicon sequencing, where we identified 53 major amplicon sequence variants (ASVs) with relative abundance >0.1% in either condition (63 unique ASVs detected overall) (Data S4B). Using the taxonomy classification analysis based on the Genome Taxonomy Database,[47,48] we classified the identified MAGs into seven bacterial classes, with Alphaproteobacteria (25 MAGs), Gammaproteobacteria (9 MAGs), and Bacteroidia

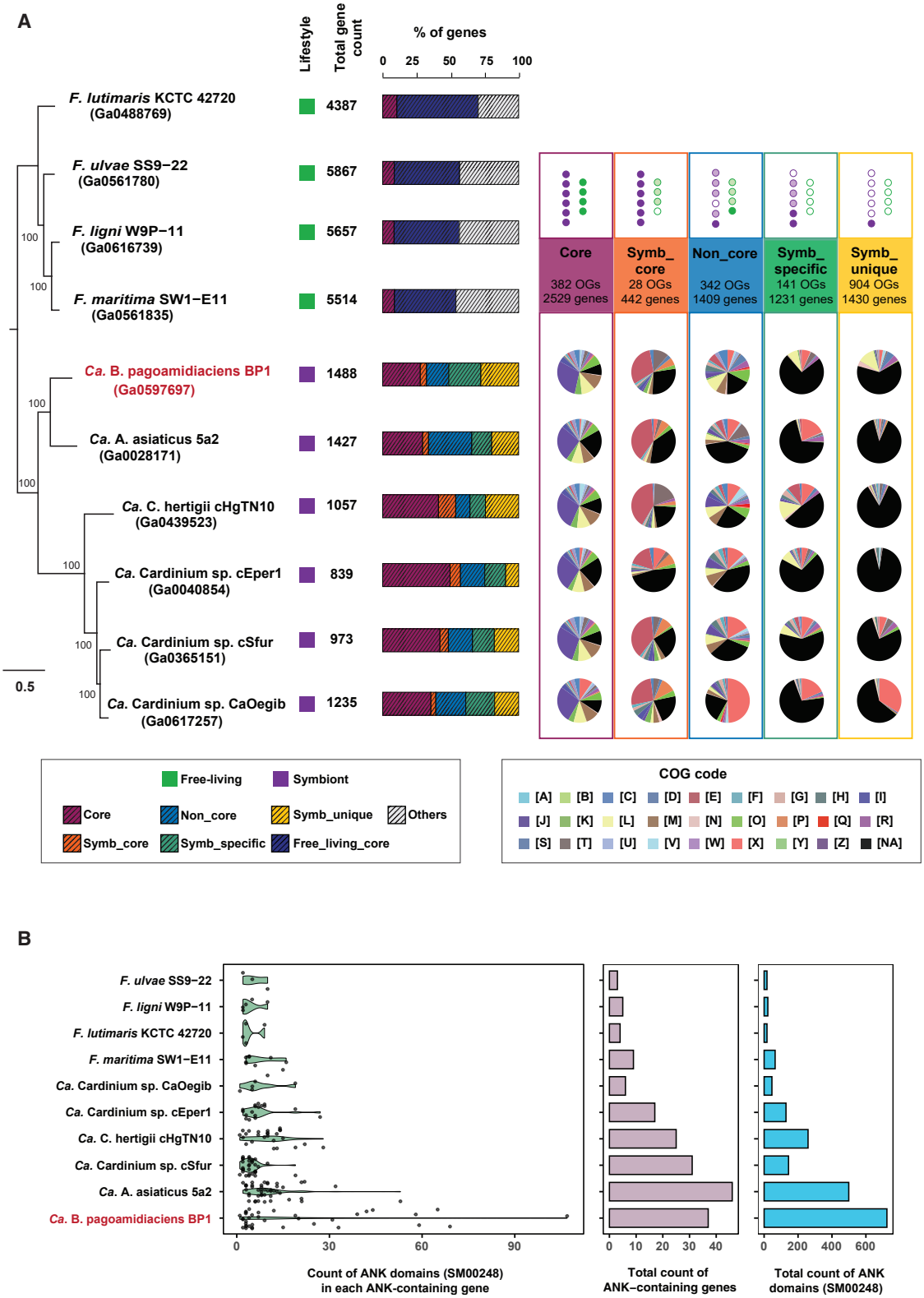


Figure 4. Comparative genomic analysis between symbiotic Amoebophilaceae and free-living *Fulvivirga* strains

(A) Whole-genome analysis of symbiotic Amoebophilaceae (purple squares) and free-living *Fulvivirga* strains (green squares) (Data S2). The phylogenetic tree was constructed using an alignment of 288 single-copy genes, identified by OrthoFinder.[38] Bar graphs indicate the percent of genes in each OG set. OG sets are illustrated by purple dots for symbiotic Amoebophilacea and green-colored dots for free-living *Fluvivirga* strains, where solid dots indicate the presence of an OG,

(legend continued on next page)

(9 MAGs) as the dominant ones. We then computed the relative abundance of these bacterial classes, using either the short-read metagenomic sequencing data or the 16S rRNA gene amplicon sequencing data, and obtained consistent results (Figure 5B) (see Data S4C and Figure S4 for microbiome profiling results at the different taxonomic levels).

We found that the overall class-level microbial community composition is similar under the PA+ and PA− conditions, where the combined abundance of Alphaproteobacteria, Gammaproteobacteria, and Bacteroidia accounts for over 96% of the total. While metagenomic and 16S rRNA gene amplicon sequencing provide similar descriptions of the microbial community associated with *Bryopsis* sp. ASF14, metatranscriptomic analysis of the same microbial community reveals a slightly different picture (Figure 5B). The class Alphaproteobacteria, despite being the most dominant on the metagenomic level (80.7% and 57.1% relative abundance for 16S rRNA gene amplicon sequencing; 79.8% and 56.9% relative abundance for metagenomic data), only accounts for 30.6% and 26.6% of the total mapped metatranscriptomic reads in the PA+ and PA− samples, respectively. On the other hand, the class Bacteroidia (10.4% and 28.4% relative abundance for 16S rRNA gene amplicon sequencing; 10.9% and 21.6% relative abundance for metagenomic data) accounts for 56.8% and 48.2% of the total mapped metatranscriptomic reads in the PA+ and PA− samples, respectively, making it the most transcriptionally active bacterial class under both conditions. Taken together, our results show that *Bryopsis* sp. ASF14 harbors a moderately diverse microbial community that is relatively stable under both cultivation conditions and that certain members of this community are much more transcriptionally active than expected from their relative abundance.

Next, we focused our analysis on *Ca.* B. pagoamidifaciens BP1 as a member of the *Bryopsis* sp. ASF14 microbiome, starting with the PA+ condition. *Ca.* B. pagoamidifaciens BP1 itself is among the top-ranked bacteria by abundance in the PA+ condition. The relative abundance of *Ca.* B. pagoamidifaciens BP1 ranks 12th among all ASVs in the PA+ sample (1.2%, using the 16S rRNA gene amplicon sequencing data) and 5th among all MAGs (3.8%, using metagenomic data) (Figure 5C). Strikingly, *Ca.* B. pagoamidifaciens BP1 accounts for 26.0% of total mapped metatranscriptomic reads (∼7 times higher than its metagenomic abundance), making it the most transcriptionally active bacterium in the entire microbial community (Data S4A).

Our initial analysis of the PA+ and PA− algal samples showed that both abundance and expression of *pag* are highly reduced in the PA− sample (Figures S1D and S1E). Not surprisingly, we found that the observed reduction of *pag* is consistent with an overall reduction of *Ca.* B. pagoamidifaciens BP1. The relative abundance of *Ca.* B. pagoamidifaciens BP1 decreased from 1.21% of all bacterial ASVs in the PA+ sample to 0.02% in the PA− sample (using 16S rRNA gene amplicon sequencing data) and from 3.8% to 0.04% using metagenomic data. Along with the reduction in its abundance, the transcripts of *Ca.* B. pagoamidifaciens BP1 also decreased ∼20 times in the PA− sample, from 26.0% (PA+) to 1.2% (PA−) of total mapped metatranscriptomic reads (Data S4A). Finally, to quantitatively determine whether *Ca.* B. pagoamidifaciens BP1 *pag* expression level varies between the two conditions, we calculated the percent of *Ca.* B. pagoamidifaciens BP1 transcripts contributed by *pag* under the two conditions (*pag*'s read counts/*Ca.* B. pagoamidifaciens BP1's whole genome read counts in the PA+ and PA− metatranscriptomes X 100). We found that *pag*'s transcriptional contribution accounts for 5.6% of the *Ca.* B. pagoamidifaciens BP1 genome's transcriptional activity in the PA+ sample, as opposed to 3.3% in the PA− sample. This small decrease in *pag*'s transcriptional contribution in the PA− samples (1.7-fold) is dwarfed by the overall decrease in *Ca.* B. pagoamidifaciens BP1's transcripts (∼20-fold) and metagenomic abundance (∼100-fold) (Data S4D). Combined, our analysis suggests that the decrease in *pag*'s abundance and metatranscriptomic counts in the PA− sample is a mere reflection of the decreased abundance of *Ca.* B. pagoamidifaciens BP1, as opposed to a transcriptional regulation outcome.

Varying symbiont load controls pagoamide A levels

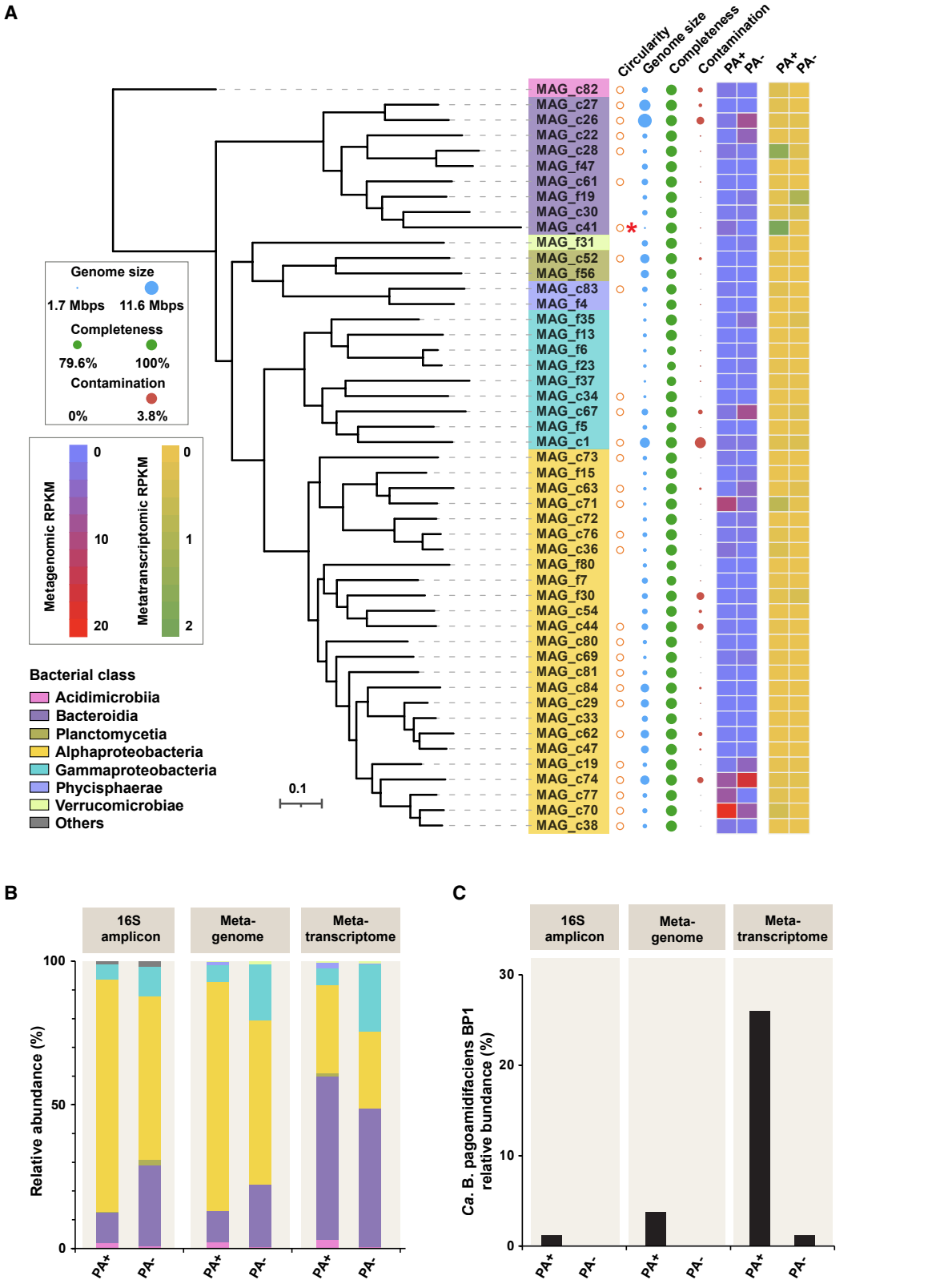
Temperature control of symbiont load is an interesting finding, especially since we did not observe any noticeable difference in algal morphology, biomass, or growth between the two conditions. Next, we wondered how long it takes to cause a significant drop or increase in the producing symbiont's load and, in turn, in pagoamide A's levels, and whether these changes are readily reversible. To answer these questions, we set up a 63-day crossover growth experiment as follows: we started with 42 identical subcultures of PA+ algae (50 mL scale) and cultivated half of which at 20°C and half at 27°C for 28 days. At day 28, half of the 20°C cultures were switched to 27°C, and half of the 27°C cultures were switched to 20°C, and all cultures were allowed to grow for 35 additional days. We collected longitudinal cultures periodically and quantitatively monitored the absolute load of *Ca.* B. pagoamidifaciens BP1 and yield of pagoamide A using digital droplet PCR (ddPCR) and quantitative metabolomics, respectively (STAR Methods; Figure 6A).

Four key findings emerged from this experiment (Figure 6B). First, in the 27°C stable cultures, the bacterial load of *Ca.* B. pagoamidifaciens BP1 continued to increase and reached up to

hollow dots indicate the absence of an OG, and low-opacity dots indicate either the presence or absence of an OG. Core, OG is found in all analyzed genomes; symbiont_core, OG is found in all symbiotic genomes and could be present or absent in free-living ones; symbiont_specific, OG is found in multiple symbiotic genomes but is not present in free-living ones; symbiont_unique, OG is found uniquely in one symbiotic genome; non_core, OG doesn't meet any of the above scenarios in symbiotic genomes; free-living core, OG is found in all free-living genomes and could be present or absent in symbiotic ones; others, OG doesn't belong to "core" or "free-living core" in free-living genomes. Pie charts indicate the COG assignment of the genes in each OG set for symbiotic Amoebophilaceae strains.

(B) Comparison of genes encoding for ANK-containing proteins in Ameobophilaceae symbionts and free-living *Fulvivirga* strains. The ANK domain was identified using InterProScan5 with domain SMART accession #SM00248.[39] *Ca.* B. pagoamidifaciens BP1 is highlighted in red.

See also Data S2.



(legend on next page)

1,903 ± 215 copies/ng DNA. On the other hand, in the 20°C stable cultures, the load of *Ca.* B. pagoamidifaciens BP1 gradually decreased until it reached 10 ± 3 copies/ng DNA. Second, in the cultures that were switched from 27°C to 20°C, the changes in *Ca.* B. pagoamidifaciens BP1's levels were typically quick: during the first switch (at day 0), *Ca.* B. pagoamidifaciens BP1's level dropped from 158 ± 31 copies/ng DNA to 30 ± 10 copies/ng DNA within 9 days, and in the crossover switch (at day 28), it again dropped from 690 ± 271 copies/ng DNA to 20 ± 5 copies/ng DNA in 10 days. Third, all changes were reversible, although recovery of *Ca.* B. pagoamidifaciens BP1 levels was slower than their drop: it took 35 days to recover from 30 ± 10 copies/ng DNA to 268 ± 200 copies/ng DNA of *Ca.* B. pagoamidifaciens BP1 when switched from 20°C back to 27°C. Finally, there was a strong dependence of the levels of pagoamide A on those of *Ca.* B. pagoamidifaciens BP1. In the 27°C stable cultures, the level of pagoamide A reached its peak (370 ± 140 ng/mg of extract) in 28 days and then slightly went down to 226 ± 103 ng/mg extract at day 63. In the 20°C stable cultures, the level of pagoamide A gradually decreased until it reached its lowest level, 2 ± 2 ng/mg of extract, at day 63. In the crossover cultures, pagoamide A's levels also followed those of *Ca.* B. pagoamidifaciens BP1, albeit, expectedly, with a slight delay: pagoamide A's levels dropped from 370 ± 140 ng/mg of extract to 123 ± 27 ng/mg of extract after 10 days of switching the cultures from 27°C to 20°C and continued to decrease until it reached 4 ± 6 ng/mg at day 63 (35 days after switching). The recovery of pagoamide A yields was much slower than their drop. It took pagoamide A's levels 20 days to slowly recover from 14 ± 4 ng/mg to 19 ± 13 ng/mg after switching the cultures from 20°C to 27°C, and the molecule continued to accumulate until it reached 131 ± 47 ng/mg at day 63 (35 days after switching). Taken together, our results suggest that *Ca.* B. pagoamidifaciens BP1 levels within the algal host are extremely sensitive to environmental temperatures, with drastic changes in its absolute load occurring within days of condition change in either direction, and that pagoamide A's levels closely follow and are fully dependent on this change in symbiont load.

Transcriptional profiles of the host alga *Bryopsis* sp. ASF14 under low and high symbiont loads

To explore whether different loads of the *Ca.* B. pagoamidifaciens BP1 symbiont (and in turn of pagoamide A) have an impact on the algal host, or, alternatively, whether exposure to different temperatures tunes the host's ability to control its symbiont load, we sought to perform comparative transcriptomic analysis between *Bryopsis* sp. ASF14 cultivated at 20°C (PA−, low symbiont load) and 27°C (PA+, high symbiont load) and identify differentially expressed genes and pathways (STAR Methods). To allow such analysis, we first generated a draft genome of *Bryopsis* sp. ASF14 from the long-read-based metagenomic assembly: 1,174 contigs, an N50 of 219,060 bps, and a total genome size of 149,703,704 bps. Our genome annotation efforts predicted a total of 14,031 genes with a mean length of 2,395 bps, 7,751 of which have a functional hit to UniProtKB/Swiss-Prot database (Figure S5B; Data S5A). A BUSCO analysis of the draft genome in protein mode indicates 90.2% completeness, suggesting that our draft genome represents most host genes.[49]

With the draft genome in hand, we isolated RNA from *Bryopsis* sp. ASF14 cultivated at 20°C and 27°C (three biological replicates each, 28 days at the set condition) and performed RNA sequencing on all samples (average 20.5 M reads, Illumina, read length 2X150 bps; Data S5B). Then, we mapped reads to the *Bryopsis* sp. ASF14 genes and performed differential expression analysis using DESeq2.[50] Using an adjusted *p* value cutoff of $\leq$0.05 (Wald test) and a fold change cutoff of $\geq$2, we found 336 genes that are significantly upregulated in the 27°C condition and 409 genes that are significantly downregulated (Figure S5C; Data S5C). Out of the 745 differentially expressed genes, only 362 had a known COG classification. The most dominant COGs included posttranslational modification, protein turnover, chaperones (COG O), amino acid transport and metabolism (COG E), and signal transduction mechanisms (COG T) in both upregulated and downregulated genes (Data S5C).

Finally, we performed a Gene Ontology (GO) enrichment analysis on differentially expressed genes to identify overrepresented biological processes, molecular functions, and cellular components using topGO.[51] We identified GO:0006508 (proteolysis in biological processes) and GO:0004252 (serine-type endopeptidase activity in molecular functions) in significantly downregulated gene sets, indicating a potential reduction of protein turnover under the 27°C cultivation condition. On the other hand, we found that GO:0030036 and GO:0006457 in biological processes (actin cytoskeleton organization and protein folding, respectively), GO:0031463 and GO:0005788 in cellular components (Cul3-RING ubiquitin ligase complex and endoplasmic reticulum lumen, respectively), and GO:0051082 and GO:0016702 in molecular functions (unfolded protein binding and oxidoreductase activity, respectively) are among the top enriched GO terms in significantly upregulated gene sets under the 27°C cultivation condition (Figure S5D; Data S5D). Taken together, we found few transcriptional changes for the host algae under the two tested

Figure 5. Bacterial composition of the *Bryopsis* sp. ASF14 microbiome

(A) Phylogenetic tree of dereplicated MAGs discovered from the *Bryopsis* sp. ASF14 microbiome, along with their characteristics (genome circularity, size, completeness, and contamination), as well as their abundance (in RPKM) in metagenomic and metatranscriptomic data from the PA+ and PA− samples. The MAG corresponding to *Ca.* B. pagoamidifaciens BP1 is marked by a red asterisk.

(B) Bacterial composition (at the class level) of the *Bryopsis* sp. ASF14 microbiome based on 16S rRNA gene amplicon sequencing and metagenomic and metatranscriptomic sequencing of the PA+ and PA− samples. Bacterial composition based on metagenomic and metatranscriptomic sequencing data was computed at the MAG level, where relative abundance was calculated as the genome size scaled read count of each MAG/the sum of genome size scaled read counts of all MAGs *100 (STAR Methods). To allow for a clear comparison between the compositional analyses resulting from the 16S rRNA gene amplicon sequencing and the metagenomic and metatranscriptomic sequencing data, taxonomical classification based on the Genome Taxonomy Database was used for all, and ASVs (from the 16S rRNA gene amplicon sequencing) that were classified to a class not represented in the MAG classes (all of which had less than 1% relative abundance) were grouped under "others" (Data S4).

See also Figure S4 and Data S4.

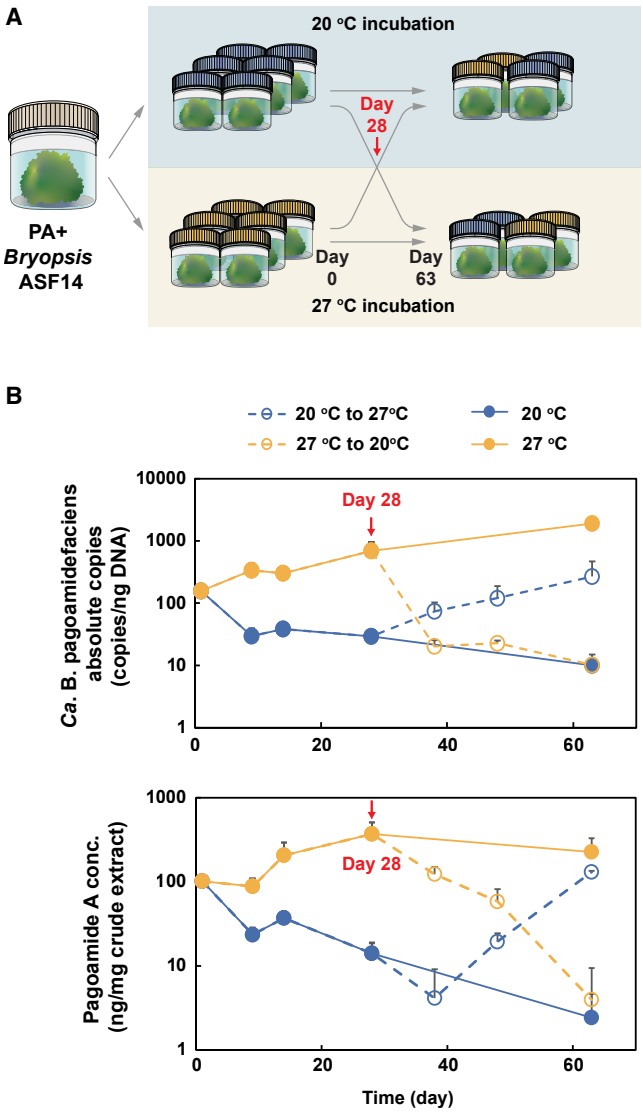


Figure 6. Temperature dependence of *Ca.* B. pagoamidifaciens BP1 and pagoamide A levels in the *Bryopsis* sp. ASF14 holobiont

(A) Experimental design of a 63-day crossover growth experiment for the *Bryopsis* sp. ASF14 holobiont.

(B) Top: absolute copies of *Ca.* B. pagoamidifaciens in algal samples, quantified using ddPCR. Bottom: absolute levels of pagoamide A in algal samples, quantified using quantitative metabolomics (HPLC-HRMS) (STAR Methods). Data points represent the mean of three biological replicates, and error bars represent the standard deviation. The red arrow indicates day 28, when conditions were shifted.

conditions despite the clear difference in environmental temperature and symbiont load.

Ca. B. pagoamidifaciens strains are widespread symbionts of marine algae

Our phylogenetic analysis of *Ca.* B. pagoamidifaciens BP1 revealed the presence of *Ca.* B. pagoamidifaciens-like symbionts (>99% identical 16S rRNA gene sequences) in environmental samples of marine algae from diverse locations (Figure 2B). This finding intrigued us to further explore the genetic variation among *Ca.* B. pagoamidifaciens symbionts obtained from different environmental sites and hosts and to determine whether or not they encode *pag*. To accomplish this goal, we obtained archived metagenomic DNA from three algal samples from which we detected *Ca.* B. pagoamidifaciens-like sequences (Figure 2B)[52] and performed short-read metagenomic sequencing on them (Data S6A). Two of the algal samples were collected from Mexico, Pacific Ocean (MX0263 and MX0019), and the third, TZ170, was collected from Tanzania, Indian Ocean. We then assembled complete algal chloroplast genomes from the metagenomic sequencing data and performed phylogenetic analysis as described above. All three chloroplast sequences were resolved in the *Bryopsis* clade of the phylogenetic tree (Figures 7A and S6). Samples MX0263 and MX0019 produced identical chloroplast genomes, suggesting that they belong to the same *Bryopsis* species. Moreover, TZ170's chloroplast is phylogenetically closer to the chloroplasts from MX0263 and MX0019 (99.7% DNA sequence identity in the *rbcL* gene, 98.89% whole chloroplast average nucleotide identity [ANI]) than the one from *Bryopsis* ASF14 (93.8% for *rbcL*, 85.56% whole chloroplast ANI) (Figure 7A).[53] Although more environmental explorations are necessary, it is interesting that we have so far only found *Ca.* B. pagoamidifaciens-like symbionts in one genus of algae, *Bryopsis* sp., and from samples that originated from tropical locations.

Next, we set out to compare the three newly identified *Ca.* B. pagoamidifaciens-like symbionts to *Ca.* B. pagoamidifaciens BP1 from *Bryopsis* sp. ASF14. The 16S rRNA gene sequence from the *Ca.* B. pagoamidifaciens-like symbiont from sample MX0263 is 100% identical to that from sample MX0019 and >99.5% identical to those from TZ170 and the original *Ca.* B. pagoamidifaciens BP1, indicating that these four organisms represent different strains of the same species. We then attempted to obtain complete genomes of the three strains from the short-read metagenomic sequencing data, but, as was observed in the case of *Ca.* B. pagoamidifaciens BP1, we only obtained highly fragmented scaffolds and therefore incomplete MAGs due to the repetitive nature of the genome. Although the genome sizes of the three recovered MAGs are only 65% to 71% of the *Ca.* B. pagoamidifaciens BP1 genome size, the sequence identity between aligned contigs of each MAG and the *Ca.* B. pagoamidifaciens BP1 genome (>1,000,000 bps of aligned sequence per MAG) is over 95% (Figure S7A; Data S6B).

In the absence of high-quality metagenomic DNA from the archived DNA samples, we could not pursue a long-read sequencing strategy as was necessary for the genome completion of *Ca.* B. pagoamidifaciens BP1. As an alternative approach, we devised a sensitive read mapping strategy for genome comparison. We directly mapped the metagenomic reads from samples MX0263, MX0019, and TZ170 to the *Ca.* B. pagoamidifaciens BP1 whole genome and computed the presence and absence of each *Ca.* B. pagoamidifaciens BP1 gene in each of the three samples (STAR Methods; Data S6C). It is important to note that this method allows us to do all comparisons with respect to the *Ca.* B. pagoamidifaciens BP1 genome but would miss genes that are unique to the new strains. We benchmarked this method by applying it to short-read metagenomic

Figure 7. Genetic variation of *Ca.* B. pagoamidifaciens strains from different geographic locations

(A) Left: a phylogenetic tree of *Ca.* B. pagoamidifaciens strains based on 187 universal prokaryotic marker genes using PhyloPhlAn.[48] Right: a phylogenetic tree of host algae based on the nucleotide sequences of their entire chloroplast genome (with several closely related sequences included). Symbiont and host trees were constructed in the same manner as in Figures 4A and S3C, respectively.

(B) Presence and absence analysis of *Ca.* B. pagoamidifaciens BP1 genes using read mapping from metagenomic sequencing data of environmental algal samples. For each *Ca.* B. pagoamidifaciens BP1 gene across the genome, an orange color indicates "presence" of the gene, and a black color indicates "absence" of the gene. White spaces between genes indicate intergenic regions in the genome (STAR Methods). Short-read metagenomic sequencing data from

(legend continued on next page)

sequencing data from the PA+ *Bryopsis* ASF14 sample and, as expected, found that 100% of the genes in the *Ca.* B. pagoamidifaciens BP1 genome are present (Figure 7B). As a negative control, we also applied the same method to short-read metagenomic sequencing data from our *Ca.* E. kahalalidifaciens-containing *Bryopsis* sp. sample SB5[19] and found that only one of the 1,488 *Ca.* B. pagoamidifaciens BP1 genes can be detected (Figure 7B). Using this approach, we found that 83.4%, 82.9%, and 85.8% of the genes in the *Ca.* B. pagoamidifaciens BP1 genome are found in samples MX0263, MX0019, and TZ170, respectively, and that 81.5% of the genes in the *Ca.* B. pagoamidifaciens BP1 genome are commonly encoded by all three *Ca.* B. pagoamidifaciens strains. These results suggest that *Ca.* B. pagoamidifaciens symbionts from different geographical locations and hosts, despite being almost identical on the 16S rRNA gene level, do harbor a variable gene repertoire.

To gain more insights into this genetic variability, we focused on *Ca.* B. pagoamidifaciens BP1's genes that are missing in the new genomes. Out of these 275 genes that do not commonly exist in all *Ca.* B. pagoamidifaciens strains, 187 of them are missing in all three new *Ca.* B. pagoamidifaciens strains, and 88 of them are variably present in the *Ca.* B. pagoamidifaciens strains. Although about half of these variable genes have no predicted function (encode for hypothetical proteins; Data S6C), we found that genes involved in eukaryotic host interactions are highly variable among the four strains. These include 25 genes encoding for ANK-containing proteins, which is the most variable gene group; five genes encoding for tetratricopeptide (TPR) repeat proteins; one gene encoding for ATP/ADP translocase (IMG Gene ID 8013675387); and two genes encoding for type IV secretory pathway VirD4 ATPases. Only 12 out of 37 ANK-containing proteins are commonly found in all four *Ca.* B. pagoamidifaciens strains (Figure 7C). Finally, the *pag* locus was also variable among the four strains: only *Ca.* B. pagoamidifaciens encodes *pag*.

To gain a better understanding of the variability around the *pag* locus, we retrieved scaffolds that map to *pag*-neighboring genes from the metagenomic assemblies of MX0263, MX0019, and TX170 and aligned them to the *Ca. B. pagoamidifaciens* BP1 genome (STAR Methods). Interestingly, all three *Ca.* B. *pagoamidifaciens* strains share perfect synteny and high sequence identity with *Ca.* B. pagoamidifaciens BP1 in regions directly upstream (21.5 kbps, 99.3% identity) of the *pag* gene cluster. They also share high sequence identity in the downstream (13.9 kbps, 99.1% identity in the aligned region) of the *pag* gene cluster except for two indel regions (Figure 7D). However, they encode different NRPSs in place of *pag*. In MX0263 and MX0019, a shorter, two-module NRPS BGC fully replaces *pag*. On the other hand, a larger NRPS BGC (incomplete thus far in our assembly) replaces *pag* in TZ170 (Figure 7D). These two NRPSs encode different modules, domains, and specificities than *pag*, indicating that they are responsible for the production of peptides of distinct structures than pagoamide A.

To elucidate the evolutionary trajectory of these NRPSs, we searched the *Ca.* B. pagoamidifaciens BP1 genome for putative horizontal gene transfer (HGT) events.[54] This analysis revealed that *pag* has been originally acquired through a potential HGT event, which is intriguing given its intracellular nature (Figure S7B). Furthermore, a DNA sequence comparison between the two NRPSs identified in the three environmental *Ca.* B. pagoamidifaciens strains and *pag* revealed a high degree of pairwise sequence identity (many regions with more than 80% nucleotide sequence identity and at least 1,000 bp of length), suggesting that the three BGCs may have originated from a common ancestor and undergone independent evolutionary divergence (Figure S7C). Conversely, *pag* shares low sequence identity with an NRPS BGC from *Ca.* Cardinium sp. CaOegib (no alignments with 80% nucleotide sequence identity and 1,000 bp of length), indicating that *pag*-like NRPSs are lineage specific within the Amoebophilaceae. Taken together, our results suggest that *Ca.* B. pagoamidifaciens strains represent a widespread lineage of algal symbionts across marine geographical landscapes and that their reduced genomes undergo genetic variations in symbiotic traits, such as the production of small molecules and microbe-host interaction genes.

DISCUSSION

Macroalgae and their microbiome function as a single ecological entity with intricate symbiotic interactions.[55,56] Algal surfaces are typically considered a hotspot for these interactions because of active nutrient exchange that promotes microbial attachment.[1] In addition, endophytic bacteria in macroalgae can also provide important ecological benefits to the host, but endosymbiosis in marine macroalgae has been mostly overlooked to date. An example for this type of interaction is between *Ca.* E. kahalalidifaciens and its *Bryopsis* algal host, where *Ca.* E. kahalalidifaciens represents an intracellular, highly functionally specialized symbiont. The genome of *Ca.* E. kahalalidifaciens possesses up to 20 NRPS BGCs and produces suites of kahalalide cocktails within the host algae, protecting it from predation.[19] Our current discovery of the *Ca.* B. pagoamidifaciens-*Bryopsis* sp. ASF14 system represents only the second identified case of a bioactive molecule produced by a yet-uncultivated macroalga-associated bacterial symbiont and differs from the *Ca.* E. kahalalidifaciens-*Bryopsis* symbiosis in several aspects.

Bryopsis sp. ASF14 serves as a positive control (all *Ca.* B. pagoamidifaciens BP1 genes are present), while that from *Bryopsis* sp. SB5 serves as a negative control (nearly all genes are absent). Genes encoding for ANK-containing proteins are indicated in blue on the *Ca.* B. pagoamidifaciens BP1 genome, while *pag* is indicated in red.

(C) Domain composition for genes encoding ANK-containing proteins in the *Ca.* B. pagoamidifaciens BP1 genome (rightmost panel), along with their expression (in RPKM) in the PA+ metatranscriptome (leftmost panel). The middle panel shows pie charts that represent the coverage of each gene (in percentage of gene length covered by metagenomic reads, green) in each of the three environmental samples harboring *Ca.* B. pagoamidifaciens strains.

(D) Genetic organization of the loci harboring *pag* or other NPRS BGCs from each of the four *Ca.* B. pagoamidifaciens strains analyzed here, along with the substrates predicted to be incorporated by each NRPS domain (X = unknown). Regions with nucleotide sequence identity >98% are shaded. PPtase: phosphopantetheinyl transferase.

See also Figures S6 and S7 and Data S6.

First, although *Ca.* B. pagoamidifaciens harbors a reduced genome similar to that of *Ca.* E. kahalalidifaciens, it does not share the same evolutionary features as *Ca.* E. kahalalidifaciens. In the latter, the kahalalide NRPS genes have undergone multiple duplication, divergence, and recombination cycles to produce considerable diversity in the genome and encoded pathways and molecules.[19] On the other hand, the *Ca.* B. pagoamidifaciens genome possesses a single NRPS BGC that is highly transcribed. Moreover, *Ca.* B. pagoamidifaciens employs unique host interaction mechanisms when compared with *Ca.* E. kahalalidifaciens, such as the large number of proteins with eukaryotic domains and ankyrin repeats. These proteins have been previously shown to mediate structural-functional relationships between intracellular bacteria and their hosts in phylogenetically related *Amoebophilus* sp. (symbionts of *Acanthamoeba*)[57] and *Cardinium* sp. (arthropod and nematode symbionts),[58] as well as in marine sponge symbionts.[43–45,59,60] Finally, while the ecological role of at least one kahalalide from *Ca.* E. kahalalidifaciens (kahalalide F) is known: chemical defense against predators,[61] the exact role of *Ca.* B. pagoamidifaciens and pagoamide A in the symbiosis remains elusive. Although pagoamide A was shown to exert some antibacterial activity against at least one bacterium,[21] suggesting a potential role in defense against pathogens or competitors, it is currently unclear how relevant this activity is to the ecology of the symbiotic system under natural environmental conditions.

Our discovery of the *Ca.* B. pagoamidifaciens-*Bryopsis* sp. system suggests that intricate symbiosis between bacteria and marine macroalgae is more common than previously appreciated. Moreover, this unique system presents a compelling example of how environmental changes—namely temperature—can have a substantial impact on the symbiosis and its functional outputs. It remains unclear, however, whether the algal host actively modulates *Ca.* B. pagoamidifaciens load and distribution. Our comparative transcriptomic analysis of *Bryopsis* sp. ASF14 cultivated under the 27°C (PA+) and 20°C (PA−) conditions did not reveal major significant changes, even in pathways related to temperature response, suggesting that this temperature difference does not trigger substantial metabolic changes in the host. Concurrently, our FISH results revealed that *Ca.* B. pagoamidifaciens can be present in high abundance in at least some algal strands under the PA+ cultivation condition.

Overall, our study describes the *Ca.* B. pagoamidifaciens-*Bryopsis* sp. ASF14 symbiotic system, its temperature dependence, and its distribution in diverse environments. Importantly, the initial findings described here pave the way for future research focused on elucidating the ecological roles of *Ca.* B. pagoamidifaciens and pagoamide A and the molecular mechanism by which the load and distribution of *Ca.* B. pagoamidifaciens are controlled.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Mohamed S. Donia (donia@princeton.edu).

Materials availability

Unique reagents generated in this study are available upon request from the lead contact.

Data and code availability

- Data S1, S2, S3, S4, S5, and S6 contain information about the sequencing data generated in this study and outputs of the various computational analyses. Sequencing data generated in this study, namely short-read metagenomic sequencing data, long-read PacBio HiFi sequencing data, and metatranscriptomic sequencing data, were deposited in the NCBI database (BioProject: PRJNA1331804). The *Ca.* B. pagoamidifaciens BP1 genome was deposited to the IMG repository, Joint Genome Institute, Department of Energy IMG (Genome: 8013675319).
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank Matthew Cahn for assistance with metagenomic and metatranscriptomic sequencing data processing, Seema Chatterjee for assistance with algal DNA extraction, Laura P. Ióca for assistance with mass spectrometry, Willem Stock for stimulating discussions about *Bryopsis*-associated bacteria, and members of the Donia laboratory for general help and useful discussions. Funding for this project was provided by the Gordon and Betty Moore Foundation (grant GBMF9199 to M.S.D., https://doi.org/10.37807/GBMF9199; grant GBMF7622 to W.H.G., https://doi.org/10.37807/GBMF7622) and NIH (grant R01GM107550 to W.H.G.).

AUTHOR CONTRIBUTIONS

J.L., M.S.D., and W.H.G. designed the project. J.L. performed all laboratory experiments and metagenomic, metatranscriptomic, and metabolomic analyses. M.S.D. aided in the metagenomic and metatranscriptomic analyses, and E.G. performed algae culture maintenance. E.G. and W.H.G. provided the initial algal culture and standard pagoamide A in pure form, and O.D.C. provided archived environmental algal samples. J.L. and M.S.D. wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

M.S.D. is a scientific co-founder and CSO at Pragma Bio. The work performed in this study is unrelated to the work performed at Pragma Bio.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Algal culturing
- METHOD DETAILS
  - Metagenomic DNA and RNA extraction
  - Metagenomic DNA, RNA and 16S rRNA amplicon sequencing
  - High molecular weight DNA extraction and PacBio HiFi sequencing
  - Illumina metagenomic, metatranscriptomic and 16S rRNA gene amplicon analyses
  - PacBio sequencing analysis and *Ca.* B. pagoamidifaciens annotation
  - Phylogenetic analysis of *Ca.* B. pagoamidifaciens and the algal host
  - Comparative genomic analysis of *Ca.* B. pagoamidifaciens and phylogenetically related bacteria
  - Fluorescence In Situ Hybridization (FISH) imaging
  - Metagenomic binning and microbial profiling
  - Cross-over growth experiment
  - Droplet Digital PCR (ddPCR) analysis
  - Chemical extraction and analysis
  - Host algal genome assembly, annotation and differential expression analysis
  - Genomic comparison between *Ca.* B. pagoamidifaciens strains

- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2025.11.023.

Received: January 20, 2025
Revised: August 14, 2025
Accepted: November 10, 2025
Published: December 16, 2025

REFERENCES

1. Ren, C.G., Liu, Z.Y., Wang, X.L., and Qin, S. (2022). The seaweed holobiont: from microecology to biotechnological applications. Microb. Biotechnol. *15*, 738–754. https://doi.org/10.1111/1751-7915.14014.
2. Croft, M.T., Lawrence, A.D., Raux-Deery, E., Warren, M.J., and Smith, A.G. (2005). Algae acquire vitamin B12 through a symbiotic relationship with bacteria. Nature *438*, 90–93. https://doi.org/10.1038/nature04056.
3. Goecke, F., Labes, A., Wiese, J., and Imhoff, J.F. (2010). Chemical interactions between marine macroalgae and bacteria. Mar. Ecol. Prog. Ser. *409*, 267–299. https://doi.org/10.3354/meps08607.
4. Singh, R.P., Mantri, V.A., Reddy, C.R.K., and Jha, B. (2011). Isolation of seaweed-associated bacteria and their morphogenesis-inducing capability in axenic cultures of the green alga Ulva fasciata. Aquat. Biol. *12*, 13–21. https://doi.org/10.3354/ab00312.
5. Provasoli, L., and Pintner, I.J. (1980). Bacteria induced polymorphism in an axenic laboratory strain of Ulva lactuca (chlorophyceae). J. Phycol. *16*, 196–201. https://doi.org/10.1111/j.1529-8817.1980.tb03019.x.
6. Alsufyani, T., Califano, G., Deicke, M., Grueneberg, J., Weiss, A., Engelen, A.H., Kwantes, M., Mohr, J.F., Ulrich, J.F., and Wichard, T. (2020). Macroalgal–bacterial interactions: identification and role of thallusin in morphogenesis of the seaweed Ulva (Chlorophyta). J. Exp. Bot. *71*, 3340–3349. https://doi.org/10.1093/jxb/eraa066.
7. Joint, I., Tait, K., Callow, M.E., Callow, J.A., Milton, D., Williams, P., and Cámara, M. (2002). Cell-to-cell communication across the prokaryote-eukaryote boundary. Science *298*, 1207. https://doi.org/10.1126/science.1077075.
8. Joint, I., Tait, K., and Wheeler, G. (2007). Cross-kingdom signalling: exploitation of bacterial quorum sensing molecules by the green seaweed Ulva. Philos. Trans. R. Soc. Lond. B Biol. Sci. *362*, 1223–1233. https://doi.org/10.1098/rstb.2007.2047.
9. Vega-Portalatino, E.J., Rosales-Cuentas, M.M., Valdiviezo-Marcelo, J., Arana-Torres, N.M., Espinoza-Espinoza, L.A., Moreno-Quispe, L.A., and Cornelio-Santiago, H.P. (2023). Antimicrobial and production of hydrolytic enzymes potentials of bacteria and fungi associated with macroalgae and their applications: a review. Front. Mar. Sci. *10*, 1174569. https://doi.org/10.3389/fmars.2023.1174569.
10. Dahms, H.U., and Dobretsov, S. (2017). Antifouling compounds from marine macroalgae. Mar. Drugs *15*, 265. https://doi.org/10.3390/md15090265.
11. Busetti, A., Maggs, C.A., and Gilmore, B.F. (2017). Marine macroalgae and their associated microbiomes as a source of antimicrobial chemical diversity. Eur. J. Phycol. *52*, 452–465. https://doi.org/10.1080/09670262.2017.1376709.
12. Biris-Dorhoi, E.-S., Michiu, D., Pop, C.R., Rotar, A.M., Tofana, M., Pop, O.L., Socaci, S.A., and Farcas, A.C. (2020). Macroalgae-a sustainable source of chemical compounds with biological activities. Nutrients *12*, 3085. https://doi.org/10.3390/nu12103085.
13. Silva, A., Silva, S.A., Carpena, M., Garcia-Oliveira, P., Gullón, P., Barroso, M.F., Prieto, M.A., and Simal-Gandara, J. (2020). Macroalgae as a source of valuable antimicrobial compounds: extraction and applications. Antibiotics (Basel) *9*, 642. https://doi.org/10.3390/antibiotics9100642.
14. Polat, S., and Ozogul, Y. (2013). Seasonal proximate and fatty acid variations of some seaweeds from the northeastern Mediterranean coast. Oceanologia *55*, 375–391. https://doi.org/10.5697/oc.55-2.375.
15. Marinho-Soriano, E., Fonseca, P.C., Carneiro, M.A., and Moreira, W.S. (2006). Seasonal variation in the chemical composition of two tropical seaweeds. Bioresour. Technol. *97*, 2402–2406. https://doi.org/10.1016/j.biortech.2005.10.014.
16. Campbell, A.H., Harder, T., Nielsen, S., Kjelleberg, S., and Steinberg, P.D. (2011). Climate change and disease: bleaching of a chemically defended seaweed. Glob. Change Biol. *17*, 2958–2970. https://doi.org/10.1111/j.1365-2486.2011.02456.x.
17. Flewelling, A.J., Ellsworth, K.T., Sanford, J., Forward, E., Johnson, J.A., and Gray, C.A. (2013). Macroalgal endophytes from the Atlantic coast of Canada: a potential source of antibiotic natural products? Microorganisms *1*, 175–187. https://doi.org/10.3390/microorganisms1010175.
18. Wiese, J., Thiel, V., Nagel, K., Staufenberger, T., and Imhoff, J.F. (2009). Diversity of antibiotic-active bacteria associated with the brown alga Laminaria saccharina from the Baltic Sea. Mar. Biotechnol. (NY) *11*, 287–300. https://doi.org/10.1007/s10126-008-9143-4.
19. Zan, J., Li, Z., Tianero, M.D., Davis, J., Hill, R.T., and Donia, M.S. (2019). A microbial factory for defensive kahalalides in a tripartite marine symbiosis. Science *364*, eaaw6732. https://doi.org/10.1126/science.aaw6732.
20. Li, Y., Yu, H.-B., Zhang, Y., Leao, T., Glukhov, E., Pierce, M.L., Zhang, C., Kim, H., Mao, H.H., Fang, F., et al. (2020). Pagoamide A, a cyclic depsipeptide isolated from a cultured marine chlorophyte, Derbesia sp., using MS/MS-based molecular networking. J. Nat. Prod. *83*, 617–625. https://doi.org/10.1021/acs.jnatprod.9b01019.
21. Wu, C.H., and Chu, J. (2021). Total synthesis and antimicrobial evaluation of pagoamide A. Front. Chem. *9*, 741290. https://doi.org/10.3389/fchem.2021.741290.
22. Arnison, P.G., Bibb, M.J., Bierbaum, G., Bowers, A.A., Bugni, T.S., Bulaj, G., Camarero, J.A., Campopiano, D.J., Challis, G.L., Clardy, J., et al. (2013). Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat. Prod. Rep. *30*, 108–160. https://doi.org/10.1039/c2np20085f.
23. Fischbach, M.A., and Walsh, C.T. (2006). Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. Chem. Rev. *106*, 3468–3496. https://doi.org/10.1021/cr0503097.
24. Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., Medema, M.H., and Weber, T. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. *47*, W81–W87. https://doi.org/10.1093/nar/gkz310.
25. Schmitz-Esser, S., Tischler, P., Arnold, R., Montanaro, J., Wagner, M., Rattei, T., and Horn, M. (2010). The genome of the Amoeba symbiont "Candidatus Amoebophilus asiaticus" reveals common mechanisms for host cell interaction among amoeba-associated bacteria. J. Bacteriol. *192*, 1045–1057. https://doi.org/10.1128/jb.01379-09.
26. Showmaker, K.C., Walden, K.K.O., Fields, C.J., Lambert, K.N., and Hudson, M.E. (2018). Genome sequence of the soybean cyst nematode (Heterodera glycines) endosymbiont "Candidatus Cardinium hertigii" strain cHgTN10. Genome Announc. *6*, e00624-18. https://doi.org/10.1128/genomeA.00624-18.
27. Zeng, Z., Fu, Y., Guo, D., Wu, Y., Ajayi, O.E., and Wu, Q. (2018). Bacterial endosymbiont Cardinium cSfur genome sequence provides insights for understanding the symbiotic relationship in Sogatella furcifera host. BMC Genomics *19*, 688. https://doi.org/10.1186/s12864-018-5078-y.
28. Halter, T., Hendrickx, F., Horn, M., and Manzano-Marín, A. (2022). A novel widespread MITE element in the repeat-rich genome of the Cardinium endosymbiont of the spider Oedothorax gibbosus. Microbiol. Spectr. *10*, e0262722. https://doi.org/10.1128/spectrum.02627-22.
29. Santos-Garcia, D., Rollat-Farnier, P.-A., Beitia, F., Zchori-Fein, E., Vavre, F., Mouton, L., Moya, A., Latorre, A., and Silva, F.J. (2014). The genome of Cardinium cBtQ1 provides insights into genome reduction, symbiont motility, and its settlement in Bemisia tabaci. Genome Biol. Evol. *6*, 1013–1030. https://doi.org/10.1093/gbe/evu077.

30. Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. *59*, 307–321. https://doi.org/10.1093/sysbio/syq010.

31. Daims, H., Brühl, A., Amann, R., Schleifer, K.H., and Wagner, M. (1999). 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. https://doi.org/10.1016/S0723-2020(99)80053-8.

32. Miravet-Verde, S., Lloréns-Rico, V., and Serrano, L. (2017). Alternative transcriptional regulation in genome-reduced bacteria. Curr. Opin. Microbiol. *39*, 89–95. https://doi.org/10.1016/j.mib.2017.10.022.

33. Halter, T., Köstlbacher, S., Collingro, A., Sixt, B.S., Tönshoff, E.R., Hendrickx, F., Kostanjšek, R., and Horn, M. (2022). Ecology and evolution of chlamydial symbionts of arthropods. ISME Commun. *2*, 45. https://doi.org/10.1038/s43705-022-00124-5.

34. Moran, N.A., and Plague, G.R. (2004). Genomic changes following host restriction in bacteria. Curr. Opin. Genet. Dev. *14*, 627–633. https://doi.org/10.1016/j.gde.2004.09.003.

35. McCutcheon, J.P., and Moran, N.A. (2011). Extreme genome reduction in symbiotic bacteria. Nat. Rev. Microbiol. *10*, 13–26. https://doi.org/10.1038/nrmicro2670.

36. Boscaro, V., Kolisko, M., Felletti, M., Vannini, C., Lynn, D.H., and Keeling, P.J. (2017). Parallel genome reduction in symbionts descended from closely related free-living bacteria. Nat. Ecol. Evol. *1*, 1160–1167. https://doi.org/10.1038/s41559-017-0237-0.

37. Schmitz-Esser, S., Linka, N., Collingro, A., Beier, C.L., Neuhaus, H.E., Wagner, M., and Horn, M. (2004). ATP/ADP translocases: a common feature of obligate intracellular amoebal symbionts related to Chlamydiae and Rickettsiae. J. Bacteriol. *186*, 683–691. https://doi.org/10.1128/jb.186.3.683-691.2004.

38. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. *20*, 238. https://doi.org/10.1186/s13059-019-1832-y.

39. Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236–1240. https://doi.org/10.1093/bioinformatics/btu031.

40. Penz, T., Horn, M., and Schmitz-Esser, S. (2010). The genome of the Amoeba symbiont "Candidatus Amoebophilus asiaticus" encodes an afp-like prophage possibly used for protein secretion. Virulence *1*, 541–545. https://doi.org/10.4161/viru.1.6.13800.

41. Böck, D., Medeiros, J.M., Tsao, H.-F., Penz, T., Weiss, G.L., Aistleitner, K., Horn, M., and Pilhofer, M. (2017). In situ architecture, function, and evolution of a contractile injection system. Science *357*, 713–717. https://doi.org/10.1126/science.aan7904.

42. Jernigan, K.K., and Bordenstein, S.R. (2014). Ankyrin domains across the Tree of Life. PeerJ *2*, e264. https://doi.org/10.7717/peerj.264.

43. Al-Khodor, S., Price, C.T., Kalia, A., and Abu Kwaik, Y. (2010). Functional diversity of ankyrin repeats in microbial proteins. Trends Microbiol. *18*, 132–139. https://doi.org/10.1016/j.tim.2009.11.004.

44. Díez-Vives, C., Moitinho-Silva, L., Nielsen, S., Reynolds, D., and Thomas, T. (2017). Expression of eukaryotic-like protein in the microbiome of sponges. Mol. Ecol. *26*, 1432–1451. https://doi.org/10.1111/mec.14003.

45. Nguyen, M.T.H.D., Liu, M., and Thomas, T. (2013). Ankyrin-repeat proteins from sponge symbionts modulate amoebal phagocytosis. Mol. Ecol. *23*, 1635–1645. https://doi.org/10.1111/mec.12384.

46. Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., and Wang, Z. (2019). MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ *7*, e7359. https://doi.org/10.7717/peerj.7359.

47. Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2019). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics *36*, 1925–1927. https://doi.org/10.1093/bioinformatics/btz848.

48. Asnicar, F., Thomas, A.M., Beghini, F., Mengoni, C., Manara, S., Manghi, P., Zhu, Q., Bolzan, M., Cumbo, F., May, U., et al. (2020). Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. Nat. Commun. *11*, 2500. https://doi.org/10.1038/s41467-020-16366-7.

49. Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., and Zdobnov, E.M. (2021). BUSCO Update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol. Biol. Evol. *38*, 4647–4654. https://doi.org/10.1093/molbev/msab199.

50. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

51. Alexa, A., and R Rahnenführer, J. (2024) topGO: Enrichment Analysis for Gene Ontology.

52. Hollants, J., Leliaert, F., Verbruggen, H., Willems, A., and De Clerck, O. (2013). Permanent residents or temporary lodgers: characterizing intracellular bacterial communities in the siphonous green alga Bryopsis. Proc. Biol. Sci. *280*, 20122659. https://doi.org/10.1098/rspb.2012.2659.

53. Lee, I., Ouk Kim, Y., Park, S.C., and Chun, J. (2016). OrthoANI: An improved algorithm and software for calculating average nucleotide identity. Int. J. Syst. Evol. Microbiol. *66*, 1100–1103. https://doi.org/10.1099/ijsem.0.000760.

54. Vernikos, G.S., and Parkhill, J. (2006). Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the Salmonella pathogenicity islands. Bioinformatics *22*, 2196–2203. https://doi.org/10.1093/bioinformatics/btl369.

55. González-Pech, R.A., Li, V.Y., Garcia, V., Boville, E., Mammone, M., Kitano, H., Ritchie, K.B., and Medina, M. (2023). The evolution, assembly, and dynamics of marine holobionts. Ann. Rev. Mar. Sci. *16*, 443–466. https://doi.org/10.1146/annurev-marine-022123-104345.

56. Egan, S., Harder, T., Burke, C., Steinberg, P., Kjelleberg, S., and Thomas, T. (2013). The seaweed holobiont: understanding seaweed–bacteria interactions. FEMS Microbiol. Rev. *37*, 462–476. https://doi.org/10.1111/1574-6976.12011.

57. Selberherr, E., Penz, T., König, L., Conrady, B., Siegl, A., Horn, M., and Schmitz-Esser, S. (2022). The life cycle-dependent transcriptional profile of the obligate intracellular amoeba symbiont Amoebophilus asiaticus. FEMS Microbiol. Ecol. *98*, fiac001. https://doi.org/10.1093/femsec/fiac001.

58. Mann, E., Stouthamer, C.M., Kelly, S.E., Dzieciol, M., Hunter, M.S., and Schmitz-Esser, S. (2017). Transcriptome sequencing reveals novel candidate genes for Cardinium hertigii-caused cytoplasmic incompatibility and host-cell interaction. mSystems *2*, e00141-17. https://doi.org/10.1128/mSystems.00141-17.

59. Lehman, S.S., Noriea, N.F., Aistleitner, K., Clark, T.R., Dooley, C.A., Nair, V., Kaur, S.J., Rahman, M.S., Gillespie, J.J., Azad, A.F., et al. (2018). The Rickettsial ankyrin repeat protein 2 is a type IV secreted effector that associates with the endoplasmic reticulum. mBio *9*, e00975-18. https://doi.org/10.1128/mBio.00975-18.

60. Jahn, M.T., Arkhipova, K., Markert, S.M., Stigloher, C., Lachnit, T., Pita, L., Kupczok, A., Ribes, M., Stengel, S.T., Rosenstiel, P., et al. (2019). A phage protein aids bacterial symbionts in Eukaryote immune evasion. Cell Host Microbe *26*, 542–550.e5. https://doi.org/10.1016/j.chom.2019.08.019.

61. Becerro, M.A., Goetz, G., Paul, V.J., and Scheuer, P.J. (2001). Chemical defenses of the sacoglossan mollusk Elysia rufescens and its host Alga Bryopsis sp. J. Chem. Ecol. *27*, 2287–2299. https://doi.org/10.1023/a:1012287105923.

62. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

63. Gabriel, L., Brůna, T., Hoff, K.J., Ebel, M., Lomsadze, A., Borodovsky, M., and Stanke, M. (2024). BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. Preprint at bioRxiv. https://doi.org/10.1101/2023.06.10.544449.

64. Chklovski, A., Parks, D.H., Woodcroft, B.J., and Tyson, G.W. (2023). CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. Nat. Methods *20*, 1203–1212. https://doi.org/10.1038/s41592-023-01940-w.

65. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645. https://doi.org/10.1101/gr.092759.109.

66. Kim, G.B., Gao, Y., Palsson, B.O., and Lee, S.Y. (2021). DeepTFactor: a deep learning-based tool for the prediction of transcription factors. Proc. Natl. Acad. Sci. USA *118*, e2021171118. https://doi.org/10.1073/pnas.2021171118.

67. Olm, M.R., Brown, C.T., Brooks, B., and Banfield, J.F. (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. *11*, 2864–2868. https://doi.org/10.1038/ismej.2017.126.

68. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

69. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682. https://doi.org/10.1038/nmeth.2019.

70. Ranallo-Benavidez, T.R., Jaron, K.S., and Schatz, M.C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat. Commun. *11*, 1432. https://doi.org/10.1038/s41467-020-14998-3.

71. Jin, J.J., Yu, W.B., Yang, J.B., Song, Y., dePamphilis, C.W., Yi, T.S., and Li, D.Z. (2020). GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. *21*, 241. https://doi.org/10.1186/s13059-020-02154-5.

72. Nurk, S., Walenz, B.P., Rhie, A., Vollger, M.R., Logsdon, G.A., Grothe, R., Miga, K.H., Eichler, E.E., Phillippy, A.M., and Koren, S. (2020). HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. Genome Res. *30*, 1291–1305. https://doi.org/10.1101/gr.263566.120.

73. Xie, Z., and Tang, H. (2017). ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. Bioinformatics *33*, 3340–3347. https://doi.org/10.1093/bioinformatics/btx433.

74. Kolmogorov, M., Bickhart, D.M., Behsaz, B., Gurevich, A., Rayko, M., Shin, S.B., Kuhn, K., Yuan, J., Polevikov, E., Smith, T.P.L., et al. (2020). metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat. Methods *17*, 1103–1110. https://doi.org/10.1038/s41592-020-00971-x.

75. Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics *27*, 863–864. https://doi.org/10.1093/bioinformatics/btr026.

76. Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol. *37*, 852–857. https://doi.org/10.1038/s41587-019-0209-9.

77. Pryszcz, L.P., and Gabaldón, T. (2016). Redundans: an assembly pipeline for highly heterozygous genomes. Nucleic Acids Res. *44*, e113. https://doi.org/10.1093/nar/gkw294.

78. Smit, A.F.A., Hubley, R., and Green, P. (2013–2015). RepeatMasker Open-4.0. http://www.repeatmasker.org.

79. Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. USA *117*, 9451–9457. https://doi.org/10.1073/pnas.1921046117.

80. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

81. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. *19*, 455–477. https://doi.org/10.1089/cmb.2012.0021.

82. Dobin, A., and Gingeras, T.R. (2015). Mapping RNA-seq reads with STAR. Curr. Protoc. Bioinform. *51*, 11.14.1–11.14.19. https://doi.org/10.1002/0471250953.bi1114s51.

83. Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. *6*, 1621–1624. https://doi.org/10.1038/ismej.2012.8.

84. Nordberg, H., Cantor, M., Dusheyko, S., Hua, S., Poliakov, A., Shabalov, I., Smirnova, T., Grigoriev, I.V., and Dubchak, I. (2014). The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. Nucleic Acids Res. *42*, D26–D31. https://doi.org/10.1093/nar/gkt1069.

85. Sayers, E.W., Bolton, E.E., Brister, J.R., Canese, K., Chan, J., Comeau, D.C., Connor, R., Funk, K., Kelly, C., Kim, S., et al. (2022). Database resources of the national center for biotechnology information. Nucleic Acids Res. *50*, D20–D26. https://doi.org/10.1093/nar/gkab1112.

86. Madeira, F., Pearce, M., Tivey, A.R.N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., and Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res. *50*, W276–W279. https://doi.org/10.1093/nar/gkac240.

87. Nurk, S., Walenz, B.P., Rhie, A., Vollger, M.R., Logsdon, G.A., Grothe, R., Miga, K.H., Eichler, E.E., Phillippy, A.M., and Koren, S. (2020). HiCanu, S.: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. Genome Res *30*, 1291–1305. https://doi.org/10.1101/gr.263566.120.

88. Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E.S., Fischer, A., Bock, R., and Greiner, S. (2017). GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Res. *45*, W6–W11. https://doi.org/10.1093/nar/gkx391.

89. Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. *47*, W59–W64. https://doi.org/10.1093/nar/gkz238.

90. Bi, G., Mao, Y., Xing, Q., and Cao, M. (2018). HomBlocks: A multiple-alignment construction pipeline for organelle phylogenomics based on locally collinear block searching. Genomics *110*, 18–22. https://doi.org/10.1016/j.ygeno.2017.08.001.

91. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.

92. Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. *49*, W293–W296. https://doi.org/10.1093/nar/gkab301.

93. Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics *27*, 764–770. https://doi.org/10.1093/bioinformatics/btr011.

94. Arimoto, A., Nishitsuji, K., Higa, Y., Arakaki, N., Hisata, K., Shinzato, C., Satoh, N., and Shoguchi, E. (2019). A siphonous macroalgal genome suggests convergent functions of homeobox genes in algae and land plants. DNA Res. *26*, 183–192. https://doi.org/10.1093/dnares/dsz002.

95. Harris, R.S. (2007). Improved Pairwise Alignment of Genomic DNA (Pennsylvania State University).

96. Darling, A.C.E., Mau, B., Blattner, F.R., and Perna, N.T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. *14*, 1394–1403. https://doi.org/10.1101/gr.2289704.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Bryopsis algae culture	This paper	N/A
Bryopsis DNA sample	Hollants et al.[52]	N/A
Chemicals, peptides, and recombinant proteins		
Paraformaldehyde	Electron Microscopy Sciences	Cat#15713
Formamide	Sigma-Aldrich	Cat#F9037-100ML
Calcofluor White Stain	Fisher Scientific	Cat#R21507
VectaShield antifade mounting medium	Fisher Scientific	Cat#NC9265087
QX200 EvaGreen Digital PCR Supermix	Bio-Rad Laboratories	Cat#1864034
Critical commercial assays		
MoBio Dneasy Power Biofilm Kit	Qiagen	Cat#24000-50
Zymo Direct-zol RNA Miniprep Kits	Zymo Research	Cat#R2051
NucleoBond HMW DNA kit	Takarabio	Cat#740160.2
Deposited data		
Sequencing data	This paper	BioProject ID PRJNA1331804
Assembled prokaryotic genome	This paper	IMG Genome ID 8013675319
Experimental models: Organisms/strains		
E. coli BL21	NEB	Cat#C2527I
Oligonucleotides		
EUB338I: 5'-GCTGCCTCCCGTAGGAGT-Cy3-3'	Daims et al.[31]	N/A
EUB338II: 5'-GCAGCCACCCGTAGGTGT-Cy3-3'	Daims et al.[31]	N/A
EUB338III: 5'-GCTGCCACCCGTAGGTGT-Cy3-3'	Daims et al.[31]	N/A
JL39-892R: 5'-GTTTCACTCTTGCGGGCGTA-6FAM-3'	this study	N/A
JL44-NON892R: 5'- GTTTCACTCTTACAGGCG TA-6FAM-3'	this study	N/A
JL36: 5'-AAGCCCTAAAGCCAGCTGTT-3'	this study	N/A
JL37: 5'-TTTGAGCACTGTTGGGGTGT-3'	this study	N/A
JL50: 5'-GGCGGCTCTCTAAGTCAGTG-3'	this study	N/A
JL39: 5'-GTTTCACTCTTGCGGGCGTA-3'	this study	N/A
Recombinant DNA		
Plasmid pjl002:pET28a with 16S rRNA gene of *Ca.* B. pagoamidifaciens	this study	N/A
Software and algorithms		
Alien_hunter 1.7	Vernikos et al.[54]	https://www.sanger.ac.uk/tool/alien_hunter/
antiSMASH 5.1.1	Blin et al.[24]	https://antismash.secondarymetabolites.org
Arivis software	Arivis AG	https://www.arivis.com/
BLAST 2.10.0	NCBI	https://blast.ncbi.nlm.nih.gov
Bowtie2 2.3.5.1	Langmead & Salzberg[62]	http://bowtie-bio.sourceforge.net/bowtie2
BUSCO 4.1.4	Simão et al.[49]	https://busco.ezlab.org
BRAKER3 3.0.8	Gabriel et al.[63]	https://github.com/Gaius-Augustus/BRAKER
CheckM2 1.0.1	Chklovski et al.[64]	https://github.com/chklovski/CheckM2
Circos v0.69-8	Krzywinski et al.[65]	http://circos.ca
DeepTFactor	Kim et al.[66]	https://bitbucket.org/kaistsystemsbiology/deeptfactor/src/master/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
DESeq2 3.21	Love et al.[50]	https://bioconductor.org/packages/DESeq2
dRep 3.4.3	Olm et al.[67]	https://github.com/MrOlm/drep
Subread 2.0.3 (featureCounts)	Liao et al.[68]	http://subread.sourceforge.net
Fiji (ImageJ distribution)	Schindelin et al.[69]	https://fiji.sc/
Geneious Prime	Geneious	https://www.geneious.com/
GenomeScope2.0	Ranallo-Benavidez et al.[70]	http://genomescope.org/genomescope2.0/
GetOrganelle 1.7.6.1	Jin et al.[71]	https://github.com/Kinggerm/GetOrganelle
GTDB-Tk v2.1.1	Chaumeil et al.[47]	https://github.com/Ecogenomics/GTDBTk
HiCanu 2.1.1	Nurk et al.[72]	https://github.com/marbl/canu
HiFi-MAG-Pipeline	PacificBiosciences	https://github.com/PacificBiosciences/pb-metagenomics-tools/blob/master/docs/Tutorial-HiFi-MAG-Pipeline.md
Illustrator 2024	Adobe Inc.	https://www.adobe.com/products/illustrator.html
InterProScan-5 5.56-89.0	Jones et al.[39]	https://github.com/ebi-pf-team/interproscan
ISEScan 1.7.2.3	Xie & Tang[73]	https://github.com/xiezhq/ISEScan
MetaBAT2	Kang et al.[46]	https://bitbucket.org/berkeleylab/metabat
metaFlye 2.8	Kolmogorov et al.[74]	https://github.com/fenderglass/Flye
OrthoFinder 3.0.1b1	Emms & Kelly[38]	https://github.com/davidemms/OrthoFinder
PhyloPhlAn3 3.0.67	Asnicar et al.[48]	https://github.com/biobakery/phylophlan
PRINSEQ lite 0.20.4B	Schmieder & Edwards[75]	http://prinseq.sourceforge.net
QIIME2 2023.7	Bolyen et al.[76]	https://qiime2.org
R version 4.4.0	R Core Team	https://www.R-project.org
Redundans 0.14a	Pryszcz & Gabaldón[77]	https://github.com/lpryszcz/redundans
RepeatMasker	Smit et al.[78]	http://www.repeatmasker.org
RepeatModeler2 2.0.3	Flynn et al.[79]	https://github.com/Dfam-consortium/RepeatModeler
SAMtools 1.18	Li et al.[80]	http://www.htslib.org/
SPAdes v3.11.0	Bankevich et al.[81]	https://github.com/ablab/spades
STAR 2.7.11b	Dobin et al.[82]	https://github.com/alexdobin/STAR
topGO (R package)	Alexa and Rahnenführer[51]	https://bioconductor.org/packages/topGO
MassHunter Quantitative Analysis B.09.00	Agilent Technologies	N/A
Other		
Algae chamber	Percival Scientific, Inc	#I22L

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Algal culturing

The pagoamide A positive algal cultures were originally maintained in modified SWBG11 media (1/2 $NaNO_3$ strength) under 27 °C and 16h/8h light/dark cycle as described previously.[20] The pagoamide A negative algal cultures were originally maintained in modified SWBG11 media (1/2 $NaNO_3$ strength) under room temperature (20 +/- 2 °C), and 16h/8h light/dark cycle. To maintain healthy algae, algal cultures were cut and transferred into fresh medium every 2 to 3 months.

METHOD DETAILS

Metagenomic DNA and RNA extraction

For total metagenomic DNA used for Illumina sequencing, algal samples were homogenized using a plastic pestle in a 1.5 ml tube and then extracted using the MoBio Power Biofilm DNA isolation kit (MoBio Laboratories, USA, now Qiagen, USA) according to the manufacturer's protocol, with slight modifications: extended vortex time (15 min) was used for better cell lysis, and double the amount of solution BF3 was used to increase the purity of the obtained DNA.

For total RNA used for metatranscriptomic sequencing, algal samples were collected and preserved in RNALater and stored in -80 °C before extraction. Total RNA was extracted using the Zymo Direct-zol RNA Miniprep Kits (Zymo Research) following the manufacturer's protocol with on-column DNase I treatment.

The three archived *Bryopsis* DNA samples were originally extracted by Hollants et al.,[52] which applied a modified CTAB protocol for DNA extraction.

Metagenomic DNA, RNA and 16S rRNA amplicon sequencing

Metagenomic DNA and total RNA samples were quantified using Qubit fluorometer (Invitrogen, CA) and their quality was determined using a Bioanalyzer (Agilent, USA). For the metagenomic sequencing, DNA samples were sheared into a mean size of ~500 bps using a Covaris E220 sonicator (Covaris, USA). The sheared DNA was used for Illumina library preparation with the PrepX DNA library kit (WaferGen, USA) using the automated Apollo 324TM NGS Library Prep System according to the manufacturer's protocol. The quality of each library was examined on Bioanalyzer and sequenced on Illumina NovaSeq SP 500nt Flowcell as paired ends reads (2 X 250 bp) following the manufacturer's manual. In total, 39.7M and 44.5 M raw paired-end reads were generated for PA+ and PA- metagenomic samples, respectively.

Libraries for the three archived DNA samples were prepared and sequenced as described above except that the DNA shearing step was eliminated. In total, 48.4M, 28.2M and 28.3M paired-end reads (2 X 250 bps) were generated for samples MX0263, MX0019 and TZ170, respectively.

Total RNA for microbial transcripts was treated with ribosomal RNA depletion using RiboZero Plant and RiboZero Bacteria kits (Illumina, USA) and polyA mRNA depletion to enrich bacterial symbionts' mRNA. Total RNA for host algae transcripts was treated with oligo-dT Dyanbeads kit (ThermoFisher, CA) for polyA mRNA enrichment to enrich for algal mRNA. The enriched mRNA samples were converted to cDNA and strand-specific RNA-seq libraries using the PrepX RNA-seq library preparation protocol on the Apollo 324 NGS Library Prep System (TakaraBio, CA). The libraries were QC-ed as described for DNA library preparation above and sequenced on Illumina NovaSeq SP 300nt Flowcell as paired-end reads (2X150 bps) following the manufacturer's manual. 16S rRNA amplicon sequencing was performed on the V4 region (~250 bps) following previously published protocol and primers using the MiSeq platform.[83]

High molecular weight DNA extraction and PacBio HiFi sequencing

To obtain high molecular weight genomic DNA (HMW gDNA) from the PA+ algae sample, a PA+ algal culture was homogenized in liquid nitrogen and then extracted using NucleoBond HMW DNA kit (Takara Bio Inc) according to the manufacturer's protocol. Precipitated HMW gDNA was resuspended in Buffer EB (Qiagen). gDNA quantity and quality were measured using Nanodrop and the size of HMW gDNA was evaluated using FEMTO Pulse (Agilent). To prepare PacBio HiFi library, gDNA was firstly sheared to an average size of about 15 Kbps using Megaruptor2 (Diagenode), and the quantity and size were checked using Qubit (Thermo Fisher) and FEMTO Pulse instruments. The DNA fragments were used for library preparation using HiFi express Template prep kit 2.0 (PacBio). The library was selected for $\geq$10 Kbps size fragments by electrophoresis using the Blue Pippin Prep system. The selected library was sequenced on a Sequel II (PacBio) instrument using Sequel II Binding kit 2.2 and Sequel II sequencing kit 2.0 for 30 hrs data collection movie time. The raw sequencing reads were analyzed with SMRT Link 10.2 using the CCS protocol with 99% predicted accuracy to generate HiFi ccs reads (a total of 1.92 M CCS reads were obtained, with an average read length of 15.7 Kbps, median HiFi read quality of Q33 and mean number of passes of 10).

Illumina metagenomic, metatranscriptomic and 16S rRNA gene amplicon analyses

Raw Illumina metagenomic and metatranscriptomic reads were filtered using PRINSEQ as previously described.[19,75] Filtered metagenomic reads (both read pairs and singletons) were assembled using SPAdes with default parameters.[81] Scaffolds of at least 5 Kbps from the Illumina metagenomic assembly were analyzed using antiSMASH5 for biosynthetic gene cluster identification.[24] For potential RiPP BGC discovery, primary amino acid sequence of pagoamide A (VCGCSTSAFVF) was used as a query in tBLASTn searches against the metagenomic assembly scaffolds with the following parameters "-evalue 1000, -word_size 3 and -seg no" to get a full list of hits. Results were further curated for sequences with exact matches to the query sequence and none were found. For computing metagenomic or metatranscriptomic abundance of specific sequences, Illumina quality-filtered reads were mapped to reference sequences using Bowtie2, and RPKM was computed.[62] 16S rRNA amplicon sequences were processed using Qiime2 as previously described.[19,76]

PacBio sequencing analysis and *Ca.* B. pagoamidifaciens annotation

PacBio HiFi reads were assembled using metaFlye with –pacbio-hifi and –meta setting.[74] Circularization of *Ca.* B. pagoamidifaciens genome was confirmed by mapping PacBio CCS reads to the genome and manually curating the edge of the contig. Circularized *Ca.* B. pagoamidifaciens genome was analyzed using antiSMASH5 for complete *pag* BGC annotation,[24] and the genome was submitted to the Integrated Microbial Genome platform (IMG, http://img.jgi.doe.gov) for whole genome annotation and analysis.[84] Transposases in *Ca.* B. pagoamidifaciens were predicted using ISEScan.[73] Ankyrin repeat (ANK) domain containing genes were identified using InterProScan5 for domain prediction (SMART accession #SM00248).[39] Transcriptional factors were predicted using DeepTFactor.[66] The abundance of each *Ca.* B. pagoamidifaciens' gene in each sample was performed by mapping filtered Illumina reads to *Ca.* B. pagoamidifaciens genome using Bowtie2 and then reads mapped to each gene were counted using featureCounts with parameter –minOverlap 10.[68] The RPKM for each gene was calculated using read count of each gene/ (gene size (Kbp) * total read per sample (in Millions). To calculate the relative abundance of each *Ca* B. pagoamidifaciens BP1 gene in metatranscriptomic data, we computed read count of each gene / total reads mapped to *Ca* B. pagoamidifaciens BP1 genome * 100. Genome features of *Ca.* B. pagoamidifaciens were plotted using Circos.[65]

Phylogenetic analysis of *Ca.* B. pagoamidifaciens and the algal host

Representative sequences for *Ca.* B. pagoamidifaciens 16S rRNA gene phylogenetic analysis were downloaded from NCBI based on the BLASTn hits obtained using the *Ca.* B. pagoamidifaciens 16S rRNA gene sequence as a query.[85] The sequences were aligned using Clustal Omega and a phylogenetic tree was constructed using PhyML in Geneious with a bootstrap value of 500.[30,86]

The host alga was previously classified as *Derbesia* sp. based on its morphological features and we reclassified it here as *Bryopsis* sp. based on the phylogeny of key marker genes and the chloroplast genome. The three marker genes, 18S rRNA gene, *tufA* and *rbcL* of the host alga were identified by using reference *Derbesia* sp. and *Bryopsis* sp. marker genes as queries in BLASTn searches against the PacBio assembly contigs of our metagenomic dataset. The closest hits of these marker genes on the NCBI database originated from *Bryopsis* sp. alga: 93% pairwise DNA sequence identity for the 18S rRNA gene (NCBI: AB840997) and 88.7% identity for *tufA* (NCBI: NC_026795) to *Bryopsis plumosa*, and 93.7% to *Bryopsis* sp. HV04063 for *rbcL* (NCBI: NC_037363). The circular genomes of the algal host's chloroplast and mitochondrion were assembled using hiCanu,[87] and annotated by the GeSeq web platform using built-in Bryopsidales algae chloroplast or mitochondrion genomes as references (Data S1).[88] Circular chloroplast and mitochondrion genome annotation was plotted via OGDRAW.[89] Complete or nearly complete chloroplast genomes used for comparison were downloaded from NCBI (Data S1), and the alignment of whole chloroplast genomes was conducted using the HomBlock pipeline.[90] The final concatenated alignment was 27,405 bp long and included major photosynthesis genes, e.g., *atpA*, *psaA*, *psaB*, *psbA*, *psbC*, *rbcL*, *tufA*. A phylogenetic tree was constructed using PhyML with 500 bootstraps.[91] The resulting phylogenetic tree from this analysis unequivocally placed the host alga in the *Bryopsis* lineage rather than the *Derbesia* lineage (Figure S3).

Comparative genomic analysis of *Ca.* B. pagoamidifaciens and phylogenetically related bacteria

Complete genomes and annotations from five bacterial symbionts (*Candidatus* Amoebophilus asiaticus 5a2, IMG genome id: 642555114; *Candidatus* Cardinium hertigii cHgTN10, IMG genome id: 2883446420; *Candidatus* Cardinium endosymbiont of *Sogatella furcifera* cSfur, IMG genome id: 3000336795; *Candidatus* Cardinium endosymbiont of *Oedothorax gibbosus*, IMG genome id: 8058534074; *Candidatus* Cardinium endosymbiont cEper1 of *Encarsia pergandiella*, IMG genome id: 2636415932) and four free-living *Fulvivirga* sp. (*Fulvivirga lutimaris* KCTC 42720, IMG genome id: 2994968153; *Fulvivirga maritima* SW1-E11, IMG genome id: 8065656811; *Fulvivirga ligni* W9P-11, IMG genome id: 8054486107; *Fulvivirga ulvae* SS9-22 IMG genome id: 8065650867) were retrieved from IMG for comparative genomic analyses (Data S2). Transposases, ANK gene and transcriptional factor analyses were performed on these genomes as described above for *Ca.* B. pagoamidifaciens.

Ortholog groups (OGs) analysis for *Ca.* B. pagoamidifaciens and the nine reference genomes listed above was performed using OrthoFinder.[38] OG sets were built based on their prevalence in these ten genomes, as described in Figure 4 and its legend. Genes in each OG were counted and their COG assignment was obtained from the IMG annotation. A phylogenetic tree of these ten genomes was generated by OrthoFinder using 288 single copy OGs with 500 bootstraps.

Fluorescence In Situ Hybridization (FISH) imaging

Fluorescence In Situ Hybridization (FISH) on *Bryopsis* sp. ASF14 was performed as described previously,[19] with some modifications described below. Probes EUB338I, II and III labeled with Cy3 (EUB338I: 5'-GCTGCCTCCCGTAGGAGT-Cy3-3'; EUB338II: 5'-GCAGCCACCCGTAGGTGT-Cy3-3'; EUB338III: 5'-GCTGCCACCCGTAGGTGT-Cy3-3') were used for targeting universal eubacterial 16S rRNA. *Ca.* B. pagoamidifaciens specific probe (JL39-892R: 5'-GTTTCACTCTTGCGGGCGTA-6FAM-3') and a negative probe (JL44-NON892R: 5'- GTTTCACTCTTACAGGCGTA-6FAM-3') with two mismatches to JL39-892R were designed in this study and labeled with 6-FAM. Probe specificity and method optimization were performed using Clone-FISH. Briefly, the full-length 16S rRNA gene of *Ca.* B. pagoamidifaciens was amplified (JL36: 5'-AAGCCCTAAAGCCAGCTGTT-3'; JL37: 5'-TTTGAGCACTGTTGGG GTGT-3') and cloned into pET28a with an IPTG-inducible T7-LacO promoter and the resulting plasmid (pJl002) was transformed into *E. coli* BL21(DE3). Cultures before and after induction were fixed and hybridized using JL39-892R or JL44-NON892R and EUB338I-III with a range of formamide concentrations (25-45%). The lowest formamide concentration (35%) at which a strong signal was obtained with JL39-892R and not JL44-NON892R was chosen for subsequent algal FISH experiments.

Bryopsis sp. ASF14 cultured under the 27 °C (PA+) condition was fixed in 4% paraformaldehyde (Electron Microscopy Sciences, USA) at room temperature for 4 hours, washed with 1X phosphate-buffered saline (PBS) three times, dehydrated through an ethanol gradient (30%, 50% and 70%) and finally preserved in 70% ethanol at -20 °C before further processing. Algal strands were incubated twice in 100% ethanol (15 minutes each) and further incubated twice in 100% methanol (15 minutes each) for tissue clearing and permeabilization. Algal strands were then gradually rehydrated (70%, 50% and 30% ethanol gradient) and washed in PBS buffer before hybridization. Hybridization was performed in a 6-well plate at 46 °C for 3.5 hours with 5 ng/ul of both universal eubacterial probes and *Ca.* B. pagoamidifaciens specific probe in hybridization buffer (0.9 M NaCl, 20 mM Tris-HCl [pH 7.2], 0.01% sodium dodecyl sulfate) containing 35% formamide. After hybridization, algal strands were washed twice at 48 °C (for 10 minutes each) using wash buffer (225 mM NaCl, 20 mM Tris-HCl [pH 7.2], 5 mM EDTA, 0.01% sodium dodecyl sulfate) and then washed twice with PBS (for 10 minutes each). Algal cell wall was counterstained using Calcofluor White Stain (Sigma Aldrich, USA) for 10 minutes and washed twice with water (5 minutes each). Algal strands were finally aligned on slides and mounted in VectaShield antifade mounting medium (Vector Labs, USA). Imaging was performed using a Nikon A1R laser scanning confocal microscope equipped with 405 nm, 488 nm and 561 nm laser lines. Z-stack images were acquired at 0.5 μm intervals using a 100X objective. Image stacks were processed and analyzed using arivis software (Carl Zeiss Microscopy, LLC, USA) and FIJI.[69]

Metagenomic binning and microbial profiling

Binning analyses were performed independently on the Illumina metagenomic assembly and the PacBio metagenomic assembly. Illumina metagenomic scaffolds from both PA+ and PA- samples were binned using MetaBAT2.[46] PacBio HiFi scaffolds were binned using the HiFi-MAG-Pipeline (https://github.com/PacificBiosciences/pb-metagenomics-tools/blob/master/docs/Tutorial-HiFi-MAG-Pipeline.md). MAGs retrieved from the Illumina and PacBio assemblies were dereplicated using dRep with the -N50W 2 parameter.[67] The quality of final dereplicated MAGs was evaluated using checkM2,[64] and their taxonomy was determined using GTDB-Tk.[47] The phylogenetic lineage assignment and phylogenetic tree generation of MAGs were conducted using PhyloPhlan3 with –diversity medium, –accurate parameters and edited in iTOL.[48,92] The RPKM of each MAG in metagenomic and metatranscriptomic samples was computed as described before. The relative abundance of MAGs in metagenomic and metatranscriptomic samples was calculated using genome size scaled read count of each MAG / sum of genome size scaled read counts from all MAGs *100.

Cross-over growth experiment

The PA+ algae were cut into small fragments and stirred to get a homogenized culture (no visible algal biomass clumps) in modified SWG11 medium. 50 mL of the homogenized culture were transferred into individual glass jars (42 glass jars in total). Half of these cultures were incubated in a chamber with 20 °C, 16h/8h light/dark cycle, ~780 lux light intensity and the other half were incubated under 27 °C, 16h/8h light/dark cycle, and ~780 lux light intensity (#I22L, Percival Scientific, Inc). After four weeks of incubation, eighteen algal cultures were moved into the opposite condition and incubated for another five weeks. The algal samples were collected periodically (days 0, 9, 14, 28, 38, 48 and 63) in triplicates (except for day 48, which was in duplicates) for DNA and chemical extraction. Briefly, a disposable inoculating loop was used to scrape the algae from the inner wall of the jar and tangle the algal filaments together, which allowed a complete collection of the entire biomass. A portion of the biomass was cut and collected for DNA extraction. The remaining algae were rinsed using Milli-Q water to remove excessive salt prior to chemical extraction. The algae samples were flash frozen at -80 °C before further analysis.

Droplet Digital PCR (ddPCR) analysis

DNA was extracted from algal samples using the MoBio Power Biofilm DNA isolation kit as described above. DNA samples were quantified using Qubit and diluted to make a concentration of 5 ng/μL. A *Ca.* B. pagoamidifaciens 16S rRNA specific primer set (JL50: 5'-GGCGGCTCTCTAAGTCAGTG-3'; JL39: 5'-GTTTCACTCTTGCGGGCGTA-3') was designed and used for ddPCR. Each 25 μL reaction contained 12.5 μL QX200 EvaGreen Digital PCR Supermix, 2.5 μL 1μM JL50 primer, 2.5 μL 1 μM JL39 primer, 1 μL 5 ng/μL DNA and 6.5 μL H_2O. The PCR mix was firstly processed on a QX200 Droplet Generator (Bio-Rad Laboratories, USA) to generate droplets. The droplet mixture was used for PCR on a thermal cycler, following the manufacturer's protocol. Briefly, the reaction sequence was as follows: 95 °C for 5 min , then 95 °C for 30 sec and 60 °C for 1 min for 40 cycles, then 4 °C for 5 min, then 90 °C for 5 min and a final hold at 4 °C. All steps were done with ramp rate of 2 °C/sec and the sample volume was set to 40 μL. The reactions were then carefully moved to the QX200 Droplet Reader for droplet reading and data analysis.

To accurately identify the true positive droplets and separate them from chimeric signals caused by non-specific binding, a positive control using a plasmid in which we cloned the full *Ca.* B. pagoamidifaciens 16S rRNA gene and a negative control using DNA from a *Ca.* B. pagoamidifaciens-free algal sample were employed. Assisted with the positive and negative controls, we could set up a signal intensity cutoff to identify and count the true positive droplets. The absolute copy number of *Ca.* B. pagoamidifaciens was computed using the ddPCR analysis software and then normalized by the total input DNA amount.

Chemical extraction and analysis

Algal biomasses used for chemical extraction were first rinsed with Milli-Q water to remove excessive salt and then lyophilized. The freeze-dried algae were then homogenized and extracted six times using 10 mL 2:1 CH_2Cl_2/methanol. The extracts were evaporated to dryness and then weighed. Extracts were dissolved in methanol and analyzed on an Agilent 1290 Infinity II system composed of an autosampler G7167B (kept at 10 °C), a multicolumn thermostat compartment G7116B (kept at 25 °C), and a high-speed pump G7120B, coupled to a quadrupole time-of-flight (QTOF) mass spectrometer (MS) detector with electrospray ionization (ESI) G6545 using data-dependent positive ionization mode. HPLC conditions were as follows: ElipsePlusC18 RRHD (2.1 × 50 mm × 1.8 μm, Agilent) column, mobile phase A (H2O + 0.1% formic acid), mobile phase B (MeCN + 0.1% formic acid), 0.4 ml/min flow rate, with a gradient of 5-95% B in 12 min and a final hold at 95% B for 2 min. The quantification of pagoamide A was done using MassHunter Quantitative Analysis B.09.00 and a standard curve of pure pagoamide A.

Host algal genome assembly, annotation and differential expression analysis

To obtain a draft genome for the host alga *Bryopsis* sp. ASF14, metagenomic contigs from the PacBio assembly were classified into a putative taxonomical origin by using BLASTn against several reference databases: 1) final MAG genomes retrieved from the binning analysis (bacterial origin); 2) host chloroplast and mitochondrion genomes assembled as explained above (algal origin); 3) NCBI RefSeq and nt databases (algal and bacterial origins based on the hit); 4) a *Bryopsis* genome reference (*Bryopsis* sp. KO, GenBank: GCA_030272585.1) (algal origin). Contigs putatively labeled as being of an algal origin were first filtered based on read coverage (10x -100x) and length (>1,000 bp), then genome redundancy was reduced using the Redundans pipeline.[77] For algal genome size estimation, Illumina short reads were decontaminated by removing reads that mapped to the final MAG and organelle genomes, followed by a k-mer analysis of the filtered reads using jellyfish[93] and genome size was estimated based on k-mer analysis by

GenomeScope.[70] Draft genome annotation was performed using the BRAKER3 pipeline.[63] First, a *de novo* repeat library was constructed using RepeatModeler2,[79] which was used by RepeatMasker to soft-mask the draft genome.[78] Next, RNA-Seq reads were aligned to the draft genome using STAR aligner to provide transcriptional evidence for the annotated genes.[82] Finally, the masked genome, RNA-Seq alignments, and a set of proteome references (*Bryopsis*_KO, GenBank: GCA_030272585.1; *Ostreobium_quekettii*, GenBank: GCA_905146915.1; *Caulerpa_lentillifera*[94]; *Chlamydomonas_reinhardtii*, GenBank: GCA_000002595.3; *Volvox_carteri*, GCA_000143455.1) were used as input for BRAKER3 to generate the final gene predictions. To obtain functional information, the resulting protein sequences were annotated by searching against the UniProtKB database using BLASTp (Figure S5; Data S5).

Total RNA for algal transcriptomic analysis was extracted and sequenced as explained above from freshly collected algal samples, which had been cultivated at the 20 °C or 27 °C conditions for 4 weeks in triplicates. RNA-seq reads were aligned to the host draft genome using the STAR aligner.[82] Gene-level read counts were generated using featureCounts.[68] Differential expression analysis between the two conditions was performed using the R package DESeq2.[50] Gene Ontology (GO) enrichment analysis of the significantly differentially expressed genes was conducted with the topGO R packages to identify overrepresented functional categories[51,77] (Figure S5; Data S5).

Genomic comparison between *Ca.* B. pagoamidifaciens strains

Raw Illumina metagenomic reads from three algal samples were processed, assembled and binned using the same methods used for strain BP1 Illumina metagenomic reads. MAG quality from the *Ca.* B. pagoamidifaciens strains was estimated using CheckM2.[64] Scaffolds from *Ca.* B. pagoamidifaciens MAGs were aligned with the *Ca.* B. pagoamidifaciens BP1 complete genome using LASTZ in Geneious,[95] and total alignment length and pairwise identity were calculated based only on individual alignments with a length larger than 5 Kbps. Raw reads from each sample were mapped to strain BP1 complete genome using bowtie2 and individual BP1 gene coverage was computed using Samtools.[80] Only genes with a coverage >=60% (i.e., more than 60% of the gene length is covered by reads) were considered as present in a sample. To benchmark the mapping method, metagenomic sequencing data from *Bryopsis* sp. ASF14 was used as a positive control and metagenomic sequencing data from *Ca.* E. kahalalidifaciens-containing *Bryopsis* sp. SB5 was used as a negative control.[19] Chloroplast genomes for the three algae were assembled using GetOrganelle,[71] and phylogenetic analysis was done as described above. Chloroplast whole genome alignment and *pag* region scaffolds alignment were performed using Mauve.[96] Putative horizontal gene transfer events in the *Ca.* B. pagoamidifaciens BP1 genome were predicted using Alien_hunter.[54] Comparison of NRPS BGCs from the discovered *Ca.* B. pagoamidifaciens genomes and one *Cardinium* genome (*Ca.* Cardinium sp. CaOegib) was performed using pairwise BLASTn and only alignments with identity >80% and length >1000 bp were used for plotting.

QUANTIFICATION AND STATISTICAL ANALYSIS

All quantifications and statistical analyses were conducted using R v4.4.0 or Excel. All quantifications and statistical analyses conducted in this study have been described in method details, figure legends or main text.

Current Biology, Volume *36*

Supplemental Information

Environmentally controlled production of pagoamide A in marine macroalgae by an intracellular bacterial symbiont

Jie Liu, Eugenia Glukhov, Olivier De Clerck, William H. Gerwick, and Mohamed S. Donia

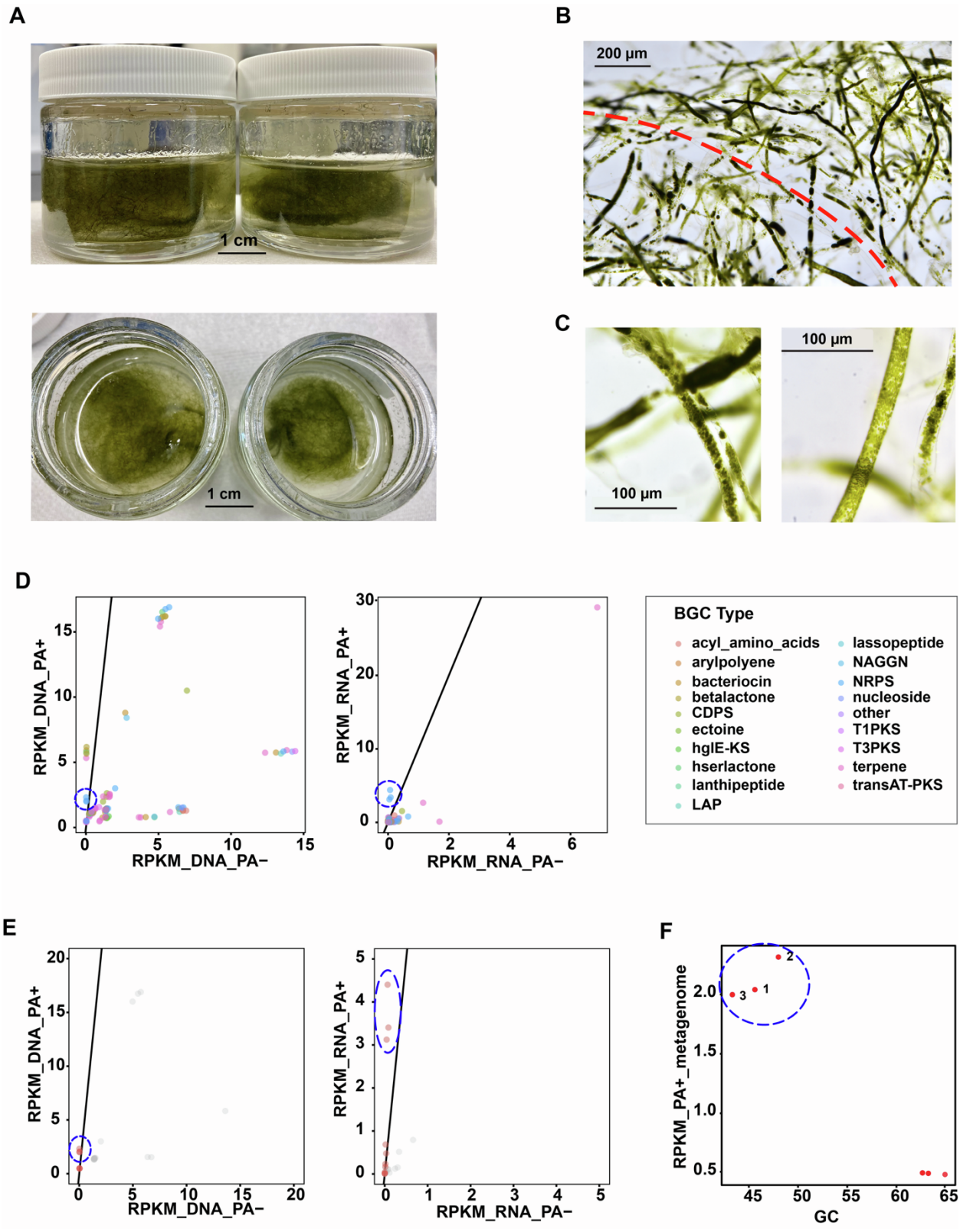
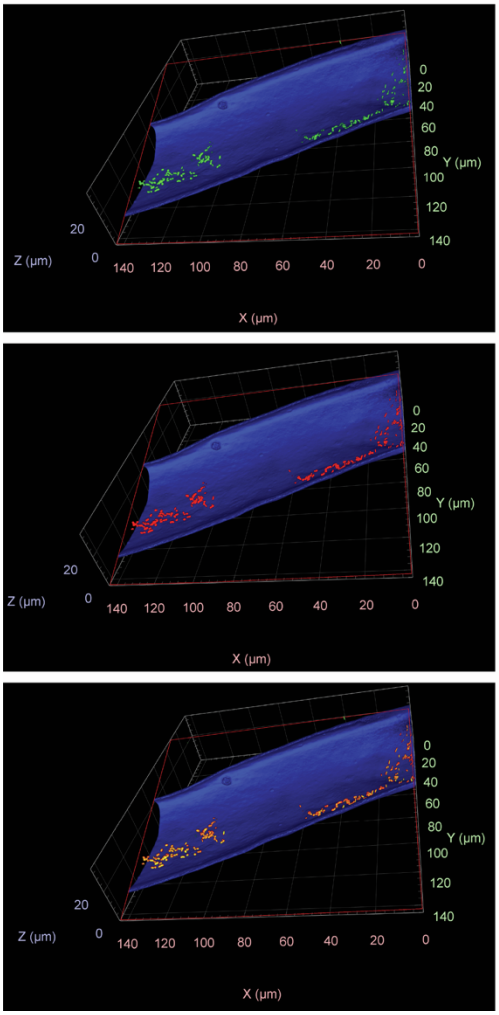

A
1 cm
1 cm
B
200 μm
C
100 μm
100 μm
D
RPKM_DNA_PA+
RPKM_DNA_PA−
RPKM_RNA_PA+
RPKM_RNA_PA−
BGC Type
acyl_amino_acids
arylpolyene
bacteriocin
betalactone
CDPS
ectoine
hglE-KS
hserlactone
lanthipeptide
LAP
lassopeptide
NAGGN
NRPS
nucleoside
other
T1PKS
T3PKS
terpene
transAT-PKS
E
RPKM_DNA_PA+
RPKM_DNA_PA−
RPKM_RNA_PA+
RPKM_RNA_PA−
F
RPKM_PA+_metagenome
GC
1
2
3

Figure S1. *Bryopsis* sp. ASF14 growth and metagenomic and metatranscriptomic abundance of BGCs from the algal microbiome under varying environmental temperatures (20 ºC or 27 ºC). Related to Figure 1. (A) Photographs of *Bryopsis* sp. ASF14 algae cultivated in glass jars at 20 ºC (left) and 27 ºC (right). (B) Micrograph of *Bryopsis* sp. ASF14 algae cultivated at 20 ºC (bottom left) and 27 ºC (top right), with a red dashed line separating the mixed cultures. (C) Micrographs of single filaments of *Bryopsis* sp. ASF14 cultivated at 20 ºC (left) and 27 ºC (right). (D) Abundance (RPKM) of identified BGCs from all biosynthetic classes in metagenomic (left) and metatranscriptomic (right) sequencing data from the PA+ and PA- algal samples. The reference line marks a 10-fold higher abundance in the PA+ sample compared to the PA- sample. (E) Abundance (RPKM) of identified BGCs from the NRPS biosynthetic class only in metagenomic (left) and metatranscriptomic (right) sequencing data from the PA+ and PA- algal samples. The reference line marks a 10-fold higher abundance in the PA+ sample compared to the PA- sample, and contigs that meet the following criterion: (PA+ RPKM / PA- RPKM >=10 or PA- RPKM = 0) are colored in red. (F) Dot plot of NRPS BGCs that pass the criterion mentioned above, showing their GC content (x axis), metagenomic abundance (RPKM) in the PA+ sample (y axis), and clearly binning them into two groups. *pag* BGC fragments are highlighted by blue circles or ellipses in all panels and numbered in F according to their location in the complete *pag* BGC (**Figure 1D**).

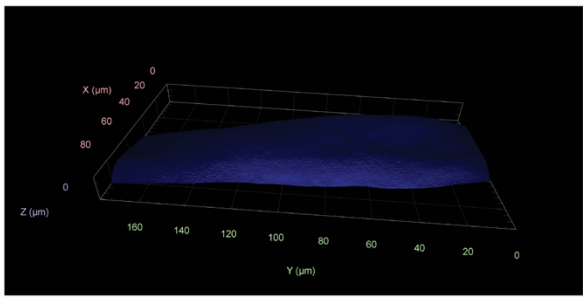
A

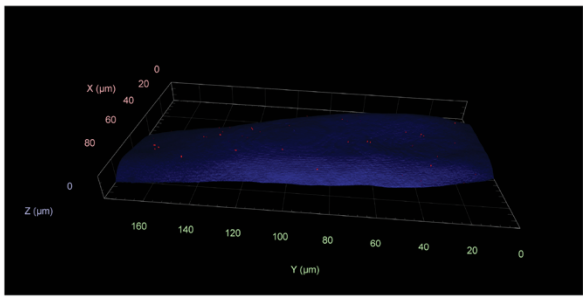
Ca. B. pagoamidifaciens BP1
Eubacteria
Cell wall



B

Ca. B. pagoamidifaciens BP1
Eubacteria
Cell wall





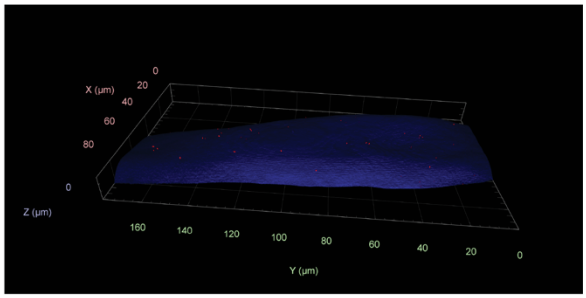
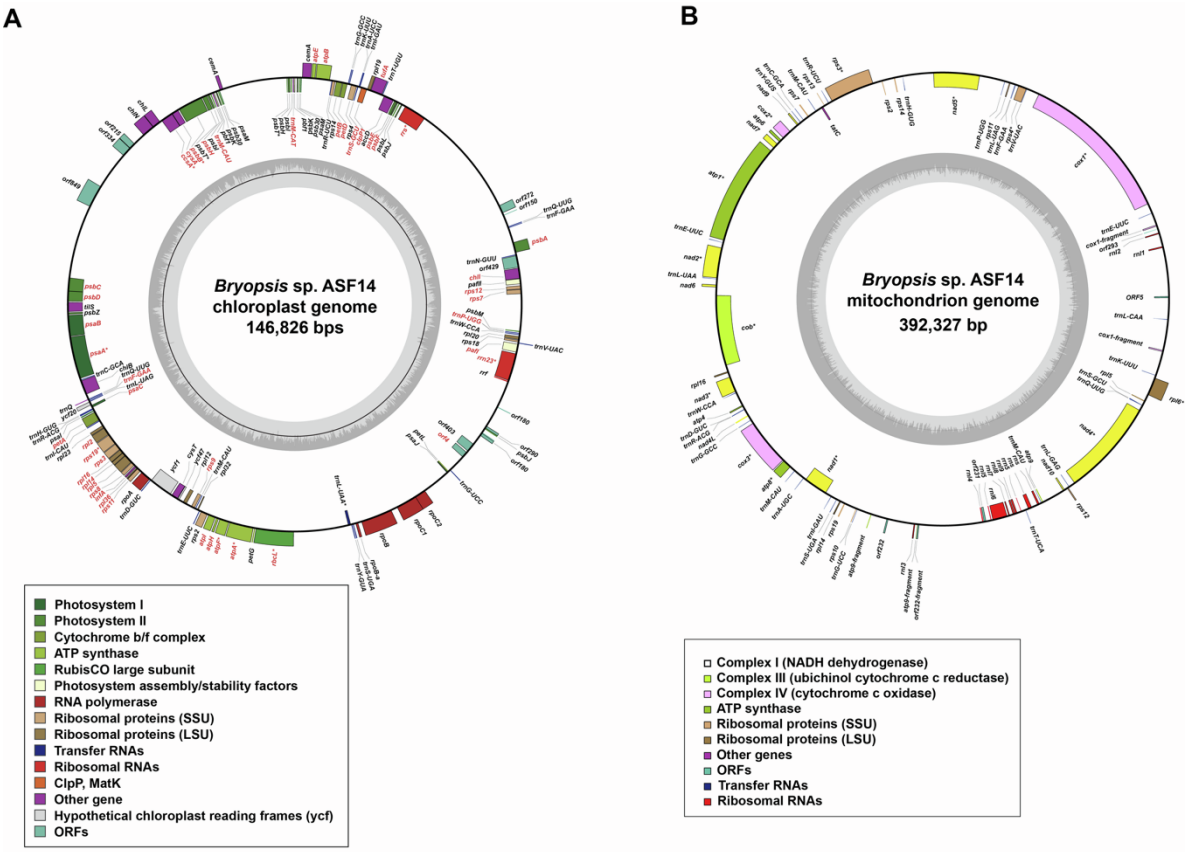


Figure S2. Three-dimensional volumetric micrographs of host algal strands. Related to Figure 3. (A) Orthogonal views of the inner face of the algal cell wall. Orthogonal projections were generated from z-stack imaging in depth of ~35 μm with a step size of 0.5 μm. (B) Orthogonal views of the outer surface of the algal cell wall. Orthogonal projections were generated from z-stack imaging in depth of ~12 μm with a step size of 0.5 μm. In both a and b, epifluorescence microscopy was performed after specimens were hybridized with universal eubacterial probes EUB338 I-III (labeled with Cy3, red) and *Ca*. B. pagoamidifaciens BP1 specific probe JL39-892R (labeled with 6-FAM, green), and counterstained with calcofluor to visualize the algal cell wall (blue). Top panel in A and B shows *Ca*. B. pagoamidifaciens BP1 specific signals detected in the green channel (488 nm), together with the cell wall stain in the blue channel (405 nm). Middle panel shows signals from the universal eubacterial probes detected in the red channel (561 nm), together with the cell wall stain in the blue channel (405 nm). Bottom panel shows a composite of the three channels (green, red, and blue).


A
Bryopsis sp. ASF14
chloroplast genome
146,826 bps
Photosystem I
Photosystem II
Cytochrome b/f complex
ATP synthase
RubisCO large subunit
Photosystem assembly/stability factors
RNA polymerase
Ribosomal proteins (SSU)
Ribosomal proteins (LSU)
Transfer RNAs
Ribosomal RNAs
ClpP, MatK
Other gene
Hypothetical chloroplast reading frames (ycf)
ORFs
B
Bryopsis sp. ASF14
mitochondrion genome
392,327 bp
Complex I (NADH dehydrogenase)
Complex III (ubichinol cytochrome c reductase)
Complex IV (cytochrome c oxidase)
ATP synthase
Ribosomal proteins (SSU)
Ribosomal proteins (LSU)
Other genes
ORFs
Transfer RNAs
Ribosomal RNAs

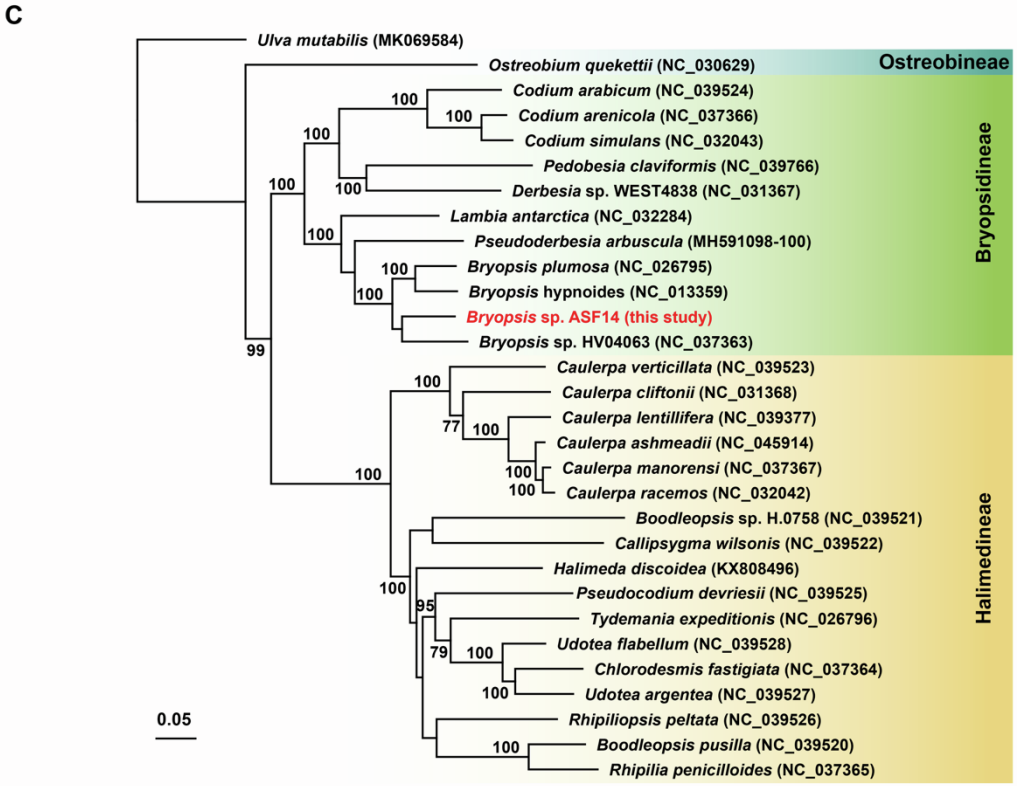
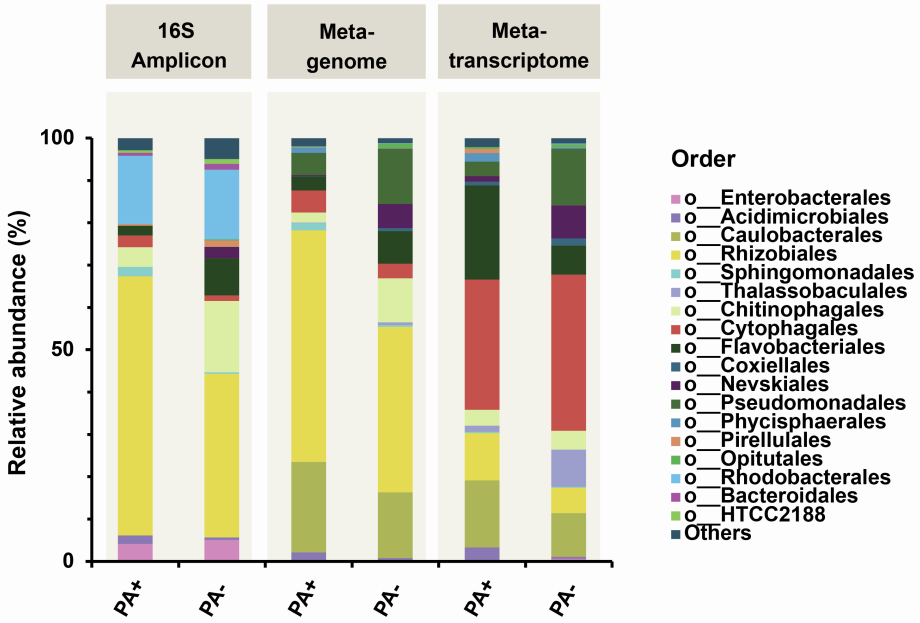

C
Ulva mutabilis (MK069584)
Ostreobium quekettii (NC_030629)
Ostreobineae
Codium arabicum (NC_039524)
Codium arenicola (NC_037366)
Codium simulans (NC_032043)
Pedobesia claviformis (NC_039766)
Derbesia sp. WEST4838 (NC_031367)
Lambia antarctica (NC_032284)
Pseudoderbesia arbuscula (MH591098-100)
Bryopsis plumosa (NC_026795)
Bryopsis hypnoides (NC_013359)
Bryopsis sp. ASF14 (this study)
Bryopsis sp. HV04063 (NC_037363)
Bryopsidineae
Caulerpa verticillata (NC_039523)
Caulerpa cliftonii (NC_031368)
Caulerpa lentillifera (NC_039377)
Caulerpa ashmeadii (NC_045914)
Caulerpa manorensi (NC_037367)
Caulerpa racemos (NC_032042)
Boodleopsis sp. H.0758 (NC_039521)
Callipsygma wilsonis (NC_039522)
Halimeda discoidea (KX808496)
Pseudocodium devriesii (NC_039525)
Tydemania expeditionis (NC_026796)
Udotea flabellum (NC_039528)
Chlorodesmis fastigiata (NC_037364)
Udotea argentea (NC_039527)
Rhipiliopsis peltata (NC_039526)
Boodleopsis pusilla (NC_039520)
Rhipilia penicilloides (NC_037365)
Halimedineae
0.05

Figure S3. Phylogenetic analysis of the host alga *Bryopsis* sp. ASF14. Related to STAR Methods. (A) Circular chloroplast genome of the host alga *Bryopsis* sp. ASF14. The GC content is indicated by the inside grey ring, where the middle back circle marks 50% GC. Genes are colored based on their general function, and their position relative to the outer ring indicates the gene orientation (forward or reverse). Names of genes included in the final alignment and used for phylogenetic analysis are colored in red. Genes with potential introns are marked with an asterisk (*) in their names. (B) Circular mitochondrion genome of the host alga *Bryopsis* sp. ASF14. The GC content is indicated by the inside grey ring, where the middle back circle marks 50% GC. Genes are colored based on their general function, and their position relative to the outer ring indicates the gene orientation (forward or reverse). Genes with potential introns are marked with an asterisk (*) in their names. (C) Phylogenetic analysis of *Bryopsis* sp. ASF14. Complete or nearly complete chloroplast genomes were aligned using HomBlock (**Data S1**), generating a 27,405 bp alignment. The phylogenetic tree was constructed using PhyML and 500 bootstraps, and bootstrap percentages are indicated next to the tree branch.

A


16S Amplicon
Meta-genome
Meta-transcriptome
Relative abundance (%)
100
50
0
PA+
PA-
Order
o__Enterobacterales
o__Acidimicrobiales
o__Caulobacterales
o__Rhizobiales
o__Sphingomonadales
o__Thalassobaculales
o__Chitinophagales
o__Cytophagales
o__Flavobacteriales
o__Coxiellales
o__Nevskiales
o__Pseudomonadales
o__Phycisphaerales
o__Pirellulales
o__Opitutales
o__Rhodobacterales
o__Bacteroidales
o__HTCC2188
Others

B

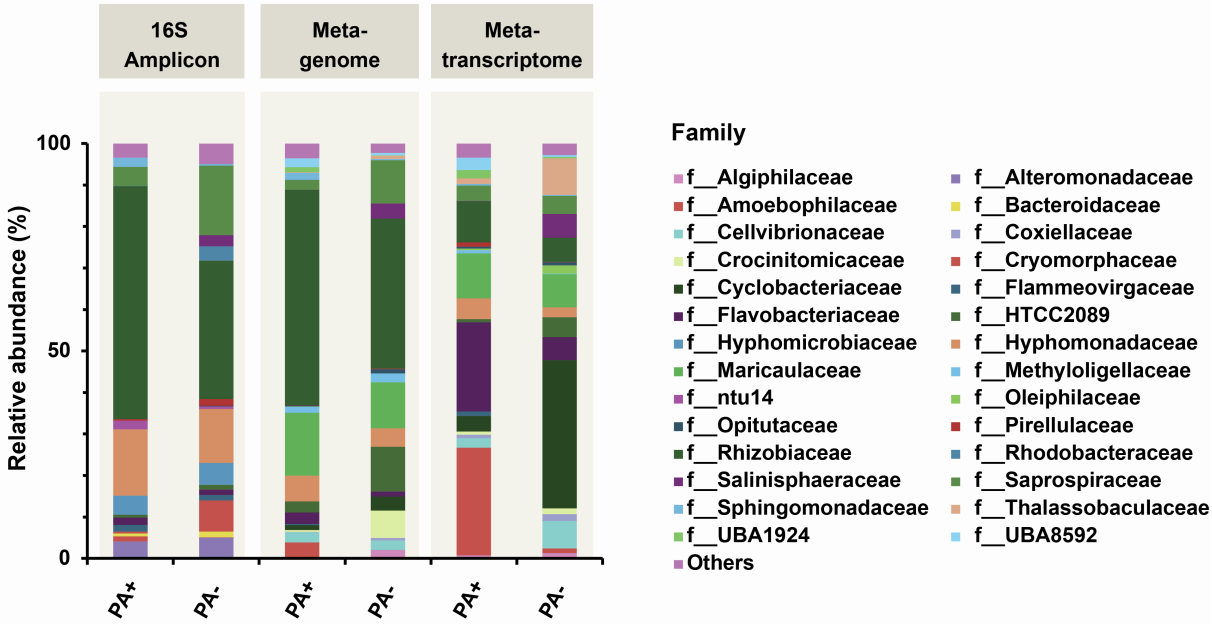
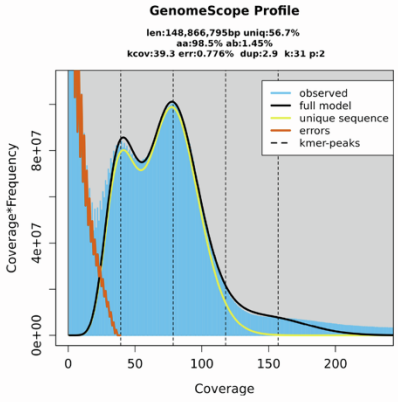

16S Amplicon
Meta-genome
Meta-transcriptome
Relative abundance (%)
100
50
0
PA+
PA-
Family
f__Algiphilaceae
f__Alteromonadaceae
f__Amoebophilaceae
f__Bacteroidaceae
f__Cellvibrionaceae
f__Coxiellaceae
f__Crocinitomicaceae
f__Cryomorphaceae
f__Cyclobacteriaceae
f__Flammeovirgaceae
f__Flavobacteriaceae
f__HTCC2089
f__Hyphomicrobiaceae
f__Hyphomonadaceae
f__Maricaulaceae
f__Methyloligellaceae
f__ntu14
f__Oleiphilaceae
f__Opitutaceae
f__Pirellulaceae
f__Rhizobiaceae
f__Rhodobacteraceae
f__Salinisphaeraceae
f__Saprospiraceae
f__Sphingomonadaceae
f__Thalassobaculaceae
f__UBA1924
f__UBA8592
Others

Figure S4. Bacterial composition of the *Bryopsis* sp. ASF14 microbiome at the order (A) and family (B) levels. Related to Figure 5. Bacterial composition of the *Bryopsis* sp. ASF14 microbiome based on 16S rRNA gene amplicon sequencing, and metagenomic and metatranscriptomic sequencing of the PA+ and PA- samples. Bacterial composition based on metagenomic and metatranscriptomic sequencing data was computed at the MAG level, where relative abundance was calculated as the genome size scaled read count of each MAG / the sum of genome size scaled read counts of all MAGs *100 (**STAR Methods**). To allow for a clear comparison between the compositional analyses resulting from the 16S rRNA gene amplicon sequencing and the metagenomic and metatranscriptomic sequencing data, taxonomical classification based on the Genome Taxonomy Database was used for all (**Data S4**). ASVs (for the 16S rRNA gene amplicon sequencing) and MAGs (for the metagenomic and metatranscriptomic sequencing) that had less than 1% relative abundance were grouped under “Others”.

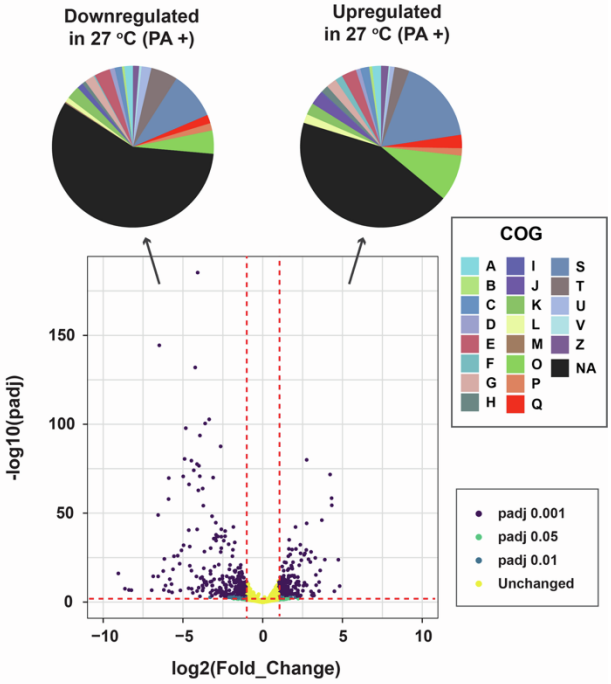
A



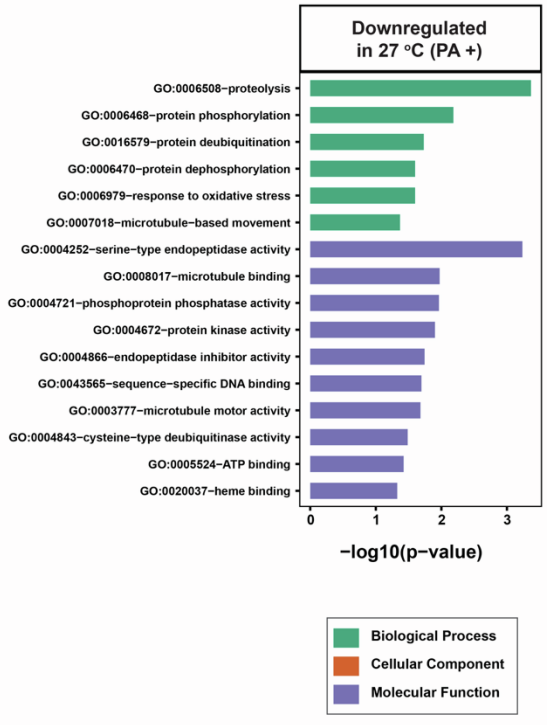
B

Bryopsis sp. ASF14 Genome	
Genome size (bp)	149,703,704
N50 (bp)	219,060
GC content (%)	49.5
Number of genes	14,031
Average gene length (bp)	2,395
Average intron per gene	4.3
Average exon length (bp)	239
Average intron length (bp)	260
Number of genes with uniprot_swissprot hit	7,751
BUSCO completeness (%)	90.2

C



D



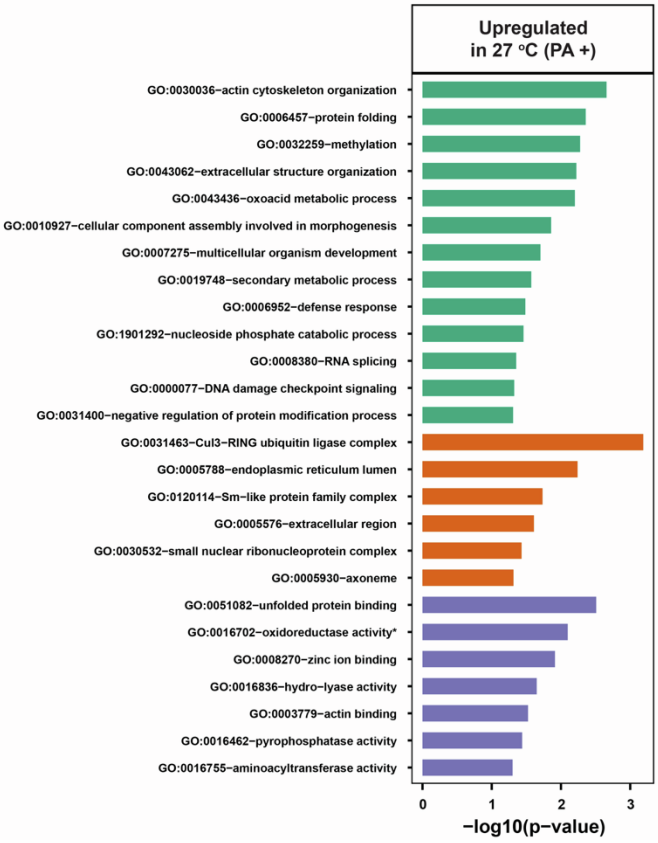


Figure S5. *Bryopsis* sp. ASF14 draft genome and comparative transcriptomic analysis. Related to STAR Methods. (A) Genome size estimation of *Bryopsis* sp. ASF14 using GenomeScope based on k-mer frequency distribution. The plot shows the observed k-mer coverage and the GenomeScope model fit with a kmer size (k) of 31 and diploid (p). Estimated genome size (len), repeat content (uniq: percent of the genome that is unique), homozygosity (aa), heterozygosity (ab), kmer coverage for heterozygous bases (kcov), sequencing error rate (err) and duplication rate (dup) are summarized in the output. (B) Summary of *Bryopsis* ASF14 draft genome assembly and annotation results (**Data S5**). (C) A volcano plot representing algal genes that are differentially expressed between the 27 °C (PA+) and 20 °C (PA-) cultivation conditions. Dashed lines indicate a cutoff of padj <=0.05 and fold change >=2. Pie charts at the top indicate the COG category classification of significantly upregulated and downregulated genes in *Bryopsis* ASF14 cultivated under the 27 °C (PA+) condition in comparison to the 20 °C (PA-) condition (**Data S5**). (D) Enriched GO terms that are significantly upregulated and downregulated (p-value <=0.05) in *Bryopsis* ASF14 cultivated under the 27 °C (PA+) condition in comparison to the 20 °C (PA-) condition (**Data S5**).

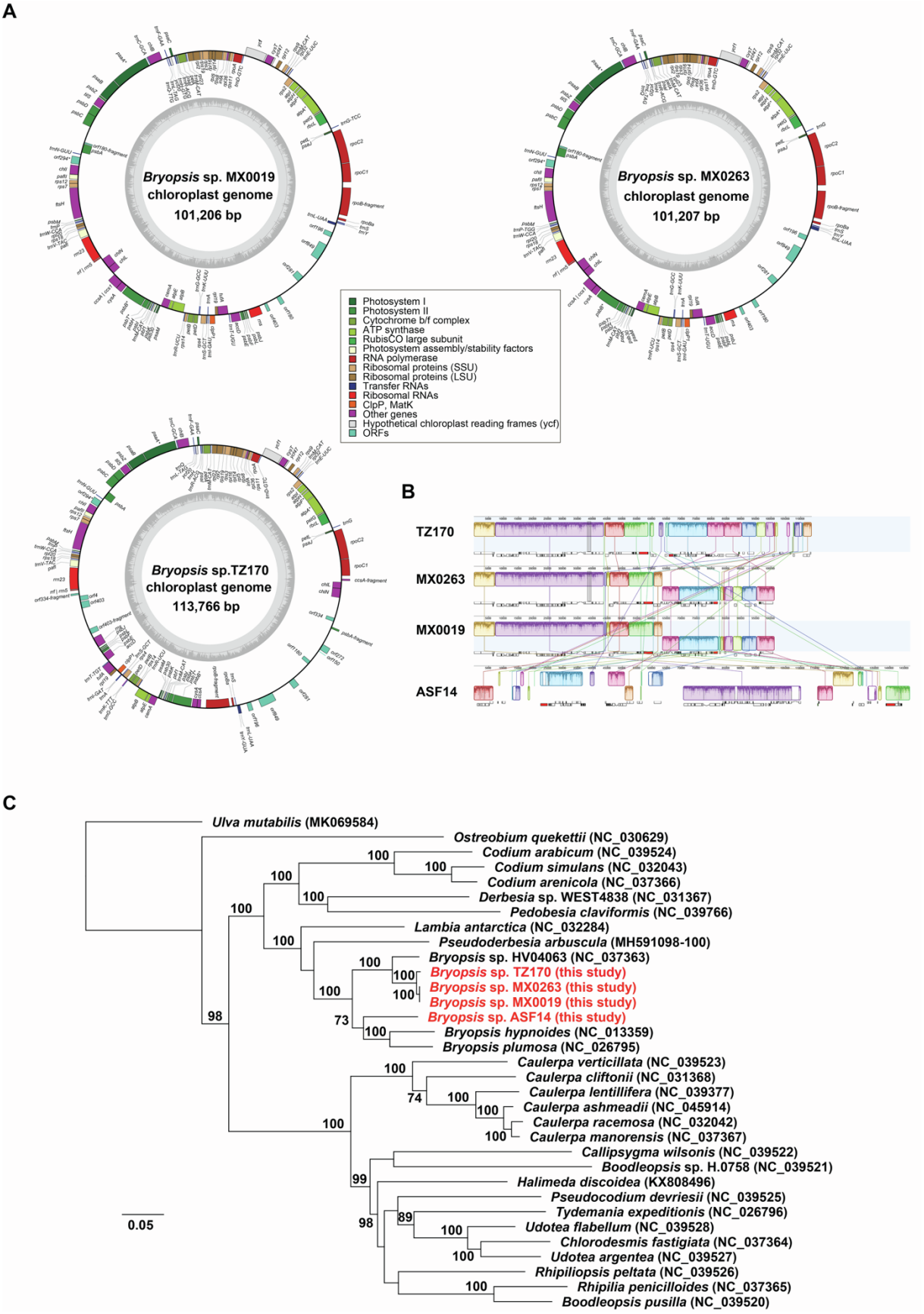

A
Bryopsis sp. MX0019
chloroplast genome
101,206 bp
Bryopsis sp. MX0263
chloroplast genome
101,207 bp
Bryopsis sp.TZ170
chloroplast genome
113,766 bp
Photosystem I
Photosystem II
Cytochrome b/f complex
ATP synthase
RubisCO large subunit
Photosystem assembly/stability factors
RNA polymerase
Ribosomal proteins (SSU)
Ribosomal proteins (LSU)
Transfer RNAs
Ribosomal RNAs
ClpP, MatK
Other genes
Hypothetical chloroplast reading frames (ycf)
ORFs
B
TZ170
MX0263
MX0019
ASF14
C
Ulva mutabilis (MK069584)
Ostreobium quekettii (NC_030629)
Codium arabicum (NC_039524)
Codium simulans (NC_032043)
Codium arenicola (NC_037366)
Derbesia sp. WEST4838 (NC_031367)
Pedobesia claviformis (NC_039766)
Lambia antarctica (NC_032284)
Pseudoderbesia arbuscula (MH591098-100)
Bryopsis sp. HV04063 (NC_037363)
Bryopsis sp. TZ170 (this study)
Bryopsis sp. MX0263 (this study)
Bryopsis sp. MX0019 (this study)
Bryopsis sp. ASF14 (this study)
Bryopsis hypnoides (NC_013359)
Bryopsis plumosa (NC_026795)
Caulerpa verticillata (NC_039523)
Caulerpa cliftonii (NC_031368)
Caulerpa lentillifera (NC_039377)
Caulerpa ashmeadii (NC_045914)
Caulerpa racemosa (NC_032042)
Caulerpa manorensis (NC_037367)
Callipsygma wilsonis (NC_039522)
Boodleopsis sp. H.0758 (NC_039521)
Halimeda discoidea (KX808496)
Pseudocodium devriesii (NC_039525)
Tydemania expeditionis (NC_026796)
Udotea flabellum (NC_039528)
Chlorodesmis fastigiata (NC_037364)
Udotea argentea (NC_039527)
Rhipiliopsis peltata (NC_039526)
Rhipilia penicilloides (NC_037365)
Boodleopsis pusilla (NC_039520)
0.05

Figure S6. Phylogenetic analysis of host algae harboring *Ca.* B. pagoamidifaciens symbionts based on their chloroplast genome sequences. Related to Figure 7. (A) Circular chloroplast genome of the host algae *Bryopsis* sp. MX0019, *Bryopsis* sp. MX0263 and *Bryopsis* sp. TZ170. The GC content is indicated by the inside grey ring, where the middle back circle marks 50% GC. Genes are colored based on their general function, and their position relative to the outer ring indicates the gene orientation (forward or reverse). Genes with potential introns are marked with an asterisk (*) in their names. (B) Whole chloroplast genome alignment for the four *Bryopsis* sp. algae harboring *Ca.* B. pagoamidifaciens symbionts. Alignment was performed using Mauve (**STAR Methods**). (C) Phylogenetic analysis of the four *Bryopsis* sp. algae harboring *Ca.* B. pagoamidifaciens symbionts (red). Complete or nearly complete chloroplast genomes were aligned using HomBlock (**Data S1**), generating a 28,248 bps alignment. The phylogenetic tree was constructed using PhyML and 500 bootstraps, and bootstrap percentages (>60%) are indicated next to the tree branch.

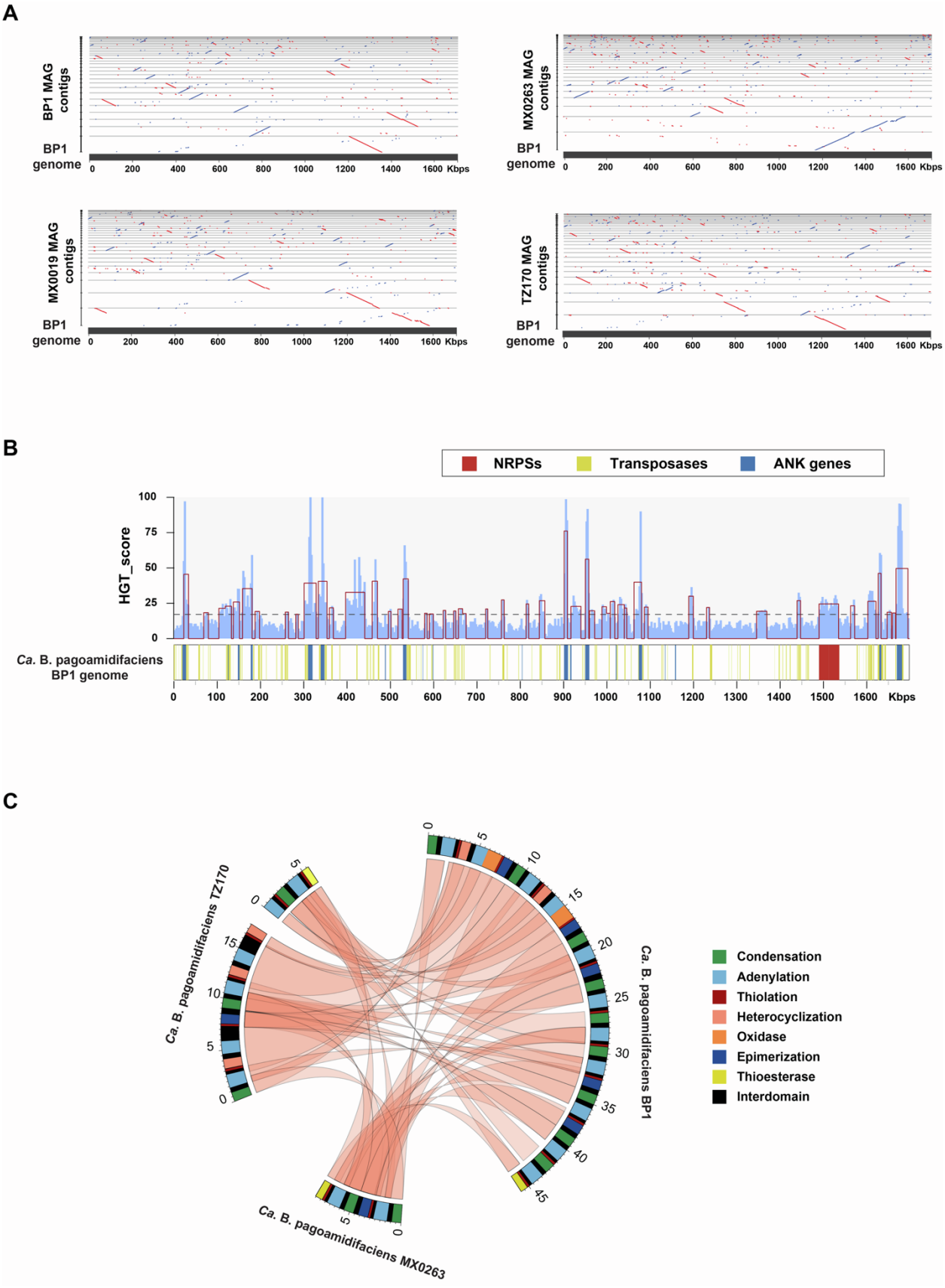

A
BP1 MAG contigs
BP1 genome
MX0263 MAG contigs
MX0019 MAG contigs
TZ170 MAG contigs
0 200 400 600 800 1000 1200 1400 1600 Kbps
B
NRPSs
Transposases
ANK genes
HGT_score
100
75
50
25
0
Ca. B. pagoamidifaciens BP1 genome
0 100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 Kbps
C
Ca. B. pagoamidifaciens TZ170
Ca. B. pagoamidifaciens MX0263
Ca. B. pagoamidifaciens BP1
Condensation
Adenylation
Thiolation
Heterocyclization
Oxidase
Epimerization
Thioesterase
Interdomain

Figure S7. Additional genomic analyses for *Ca.* B. pagoamidifaciens BP1 and related strains. Related to Figure 7. (A) Dotplot alignments (LASTZ) between contigs from various *Ca.* B. pagoamidifaciens MAGs and the complete *Ca.* B. pagoamidifaciens BP1 genome. Forward alignments are shown in blue lines, and reverse alignments in red lines. (B) Putative HGT events in the *Ca.* B. pagoamidifaciens BP1 genome. Blue bars indicate the HGT score computed in 5 Kbp windows, according to the Interpolated Variable Order Motifs (IVOMs) method used by Alien_Hunter. The score reflects the sequence composition deviation in measured windows from the genome background composition, with higher scores indicating a greater likelihood of an HGT event in the region. The horizontal reference line indicates the threshold score (17.060) for this genome, above which a putative HGT is inferred. Red lines represent putative HGT hotspots predicted using an optimized boundary localization and all regions with an HGT score below the threshold are set to 0. (C) Comparison of NRPS BGCs from different *Ca*. B. pagoamidifaciens strains. NRPS BGCs are colored based on their domain architecture. Pairwise alignments were performed between NRPS BGCs obtained from all *Ca.* B. pagoamidifaciens strains, and only alignments with >80% nucleotide sequence identity and at least 1000 bp are shown as orange ribbon links. The NRPS BGC from *Ca.* B. pagoamidifaciens strain MX0019 was not included in the analysis because it is identical to the one from strain MX0263.