Secondary metabolic symbiosis in shipworms (Teredinidae)

Authors: Marvin A. Altamia[a,b]*, Zhenjian Lin[c]*, Amaro E. Trindade-Silva[d,e], Iris Diana Uy[b,f], J. Reuben Shipway[g], Diego Veras Wilke[e], Gisela P. Concepcion[b,f], Daniel L. Distel[a], Eric W. Schmidt[c]**, Margo G. Haygood[c]**

[a]Ocean Genome Legacy Center, Department of Marine and Environmental Science, Northeastern University, Nahant, MA, USA
[b]The Marine Science Institute, University of the Philippines Diliman, Quezon City 1101, Philippines
[c]Department of Medicinal Chemistry, University of Utah
[d]Bioinformatic and Microbial Ecology Laboratory - BIOME, Federal University of Bahia, Salvador, Bahia, Brazil
[e]Drug Research and Development Center, Department of Physiology and Pharmacology, Federal University of Ceara, 60430275, Ceara, Brazil
[f]Philippine Genome Center, University of the Philippines Diliman, Quezon City 1101, Philippines
[g]Institute of Marine Science, School of Biological Sciences, University of Portsmouth, UK

*equal contribution authors
**co-corresponding authors

Significance

Shipworms play critical roles in recycling wood in the sea and in shaping mangrove habitats. Symbiotic bacteria supply the enzymes that they need for nutrition and wood degradation. Here, we show that the same nutritional symbionts also have an immense capacity to produce a multitude of diverse and likely novel bioactive secondary metabolites. The compounds likely support the ability of shipworms to degrade wood in marine environments and include a compound under investigation for its therapeutic potential. Because many of the symbionts can be cultivated, they provide a model for understanding how secondary metabolism impacts microbial symbiosis in animals.

Abstract

Shipworms, assisted by intracellular γ-proteobacteria in their gills, are the principal degraders of wood in the sea. Shipworm symbionts have been cultivated in the laboratory. The genomes of these symbionts, in addition to being replete with lytic enzymes capable of degrading wood and/or enzymes of thioautotrophic metabolism, are among the bacterial genomes richest in secondary metabolite genes. These cultivated symbionts represent the dominant species in the gills in diverse shipworm species. We investigated how the isolate secondary metabolites might impact the host animals: which bacterial pathways are present, how widely distributed they are, and how they vary. Focusing on 14 wood-eating shipworm specimens, we found between one to three major bacterial species in each gill, with each species comprising a complex mixture of closely related strains. The mixture allows the shipworm host to access a much more

complex metabolism than can be afforded by a single symbiont strain. We analyzed sequences from 22 shipworm gill metagenomes from seven shipworm species and compared them with the genomes of 23 cultivated bacterial isolates from shipworm gills. Limiting our analyses to well-characterized biosynthetic gene families, we found more than 400 polyketide, nonribosomal peptide, and related biosynthetic gene clusters, only a handful of which resemble known ones, comprising over 100 gene cluster families (GCFs). Only four GCFs are found in all specimens investigated. Several GCFs exhibited a host species-specific distribution, but most occurred stochastically. Shipworms and their symbionts thus provide a model system for understanding the role of secondary metabolism in symbioses.

Introduction

Shipworms (Family Teredinidae) are bivalve mollusks found throughout the world's oceans (1, 2). Many shipworms eat wood, assisted by cellulases from intracellular symbiotic γ-proteobacteria that inhabit their gills (Figure 1) (3-6). Others use sulfide metabolism also relying on gill-dwelling γ-proteobacteria for sulfur oxidation (7). Shipworm gill symbionts of several different species are thus essential to shipworm nutrition and survival. One of the most remarkable features of the shipworm system is that wood digestion does not take place where the bacteria are located, so that the bacterial cellulase products are transferred from the gill to a nearly sterile cecum (8), where wood digestion occurs (Figure 1) (9). This enables the host shipworms to directly consume glucose and other sugars derived from wood lignocellulose and hemicellulose, rather than the less energetic fermentation products of cellulolytic gut microbes as found in other symbioses. Shipworm symbionts are also essential for nitrogen fixation in the low-nitrogen wood environment (10). Thus, shipworms have evolved structures and mechanisms enabling bacterial metabolism to support animal host nutrition.

While in many nutritional symbioses the bacteria are difficult to cultivate, shipworm gill symbiotic γ-proteobacteria have been brought into stable culture (5, 11, 12). This led to the discovery that these bacteria are exceptional sources of secondary metabolites (13). Of bacteria with sequenced genomes, the gill symbionts *Teredinibacter turnerae* T7901 and related strains are among the richest sources of biosynthetic gene clusters (BGCs), comparable in content to famous producers of commercial importance such as *Streptomyces* spp. (12-15). This implies that shipworms might be a good source of new compounds for drug discovery. Of equal importance, the symbiotic bacteria are crucial to survival of host shipworms, and bioactive secondary metabolites might play a role in shaping those symbioses.

The genome sequence of *T. turnerae* T7901 revealed nine complex polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) BGCs (13), and more comprehensive analysis identifies up to 14 potential BGCs. One of these was shown to produce a novel catecholate siderophore, turnerbactin, which is crucial in obtaining iron and to the survival of the symbiont in nature (16). A second BGC synthesizes the borated polyketide tartrolons D/E, which are antibiotic and potently antiparasitic compounds (17). Both were detected in the extracts of shipworms, implying a potential role in producing the remarkable near sterility observed in the cecum (8). These data suggested specific roles for secondary metabolism in maintaining shipworm function.

T. turnerae T7901 is just one of multiple strains and species of γ-proteobacteria living intracellularly in various shipworm gills (3, 11), and thus these analyses just begin to describe shipworm secondary metabolism. Many shipworm species are generalists, consuming wood from a variety of sources (1, 18). Other wood-eaters, such as *Dicyathifer mannii*, *Bactronophorus thoracites*, and *Neoteredo reynei*, are specialists that live in the submerged branches, trunks and rhizomes of mangroves (19, 20). There, they play an important role in ecological processes in mangrove ecosystems, i.e. transferring large amount of carbon fixed by mangroves to the marine environment (18). Several shipworm species, such as *Kuphus*

polythalamius, live in other substrates. *K. polythalamius* often is found in sediment habitats (as well as in wood) where its gill symbionts are crucial to sulfide oxidation and carry out carbon fixation (7). *K. polythalamius* lacks significant amounts of cellulolytic symbionts such as *T. turnerae*, and instead contains *Thiosocius teredinicola*, which oxidizes sulfide and generates energy for the host (21). Other shipworms are found in solid rock and in seagrass (22, 23). Thus, gill symbionts vary, but in all cases the symbionts appear to be essential to the survival of shipworms.

While the potential of *T. turnerae* as an unexplored producer of secondary metabolites has been described (13, 15), the capacity of other shipworm symbionts is still largely unknown. Moreover, several pieces of data indicate that the BGCs found in cultivated isolates might also be found in shipworm gills, but their presence, distribution and variability in nature are unknown. Previous data include the detection of tartrolons and turnerbactins and their BGCs in shipworms (16, 17); an investigation of four isolate genomes and one metagenome that observed shared pathways (24); also an exploratory investigation of the metagenome of *N. reynei* gills and digestive tract led to the detection of known *T. turnerae* BGCs as well as novel clusters (25). These findings left major questions about the origin, abundance, variability, distribution, and potential roles of shipworm secondary metabolites. Here, we use a comparative metagenomics approach to answer these questions.

Results and Discussion

Genomic data from shipworms and their symbiotic bacteria. Cellulolytic *T. turnerae* is relatively rich in secondary metabolism in comparison to sulfide-oxidizing *T. teredinicola*. Therefore, for this analysis we selected six species of wood-eating shipworms, comparing these to a seventh sulfide-oxidizing species, *K. polythalamius*. We compared gill metagenomes from 22 specimens comprising seven animal species with the genomes of 23 cultivated bacteria isolated from shipworms (Table S1).

Of the animals obtained, we analyzed three specimens each of *Bactronophorus thoracites*, *Kuphus* spp., *Neoteredo reynei*, and *Teredo* sp., two specimens of *Bankia* sp., and five specimens of *Bankia setacea*. These animals were divided into three geographical regions (Figure 1): the Philippines (*B. thoracites* and *D. mannii* from Infanta, Quezon; *Kuphus* spp. from Mindanao and Mabini); Brazil (*N. reynei* from Rio de Janeiro, *Teredo* sp., and *Bankia* sp. from Ceará); and the United States (*B. setacea*). The purpose of sampling this range was to determine whether there are any geographical differences in gill symbionts. Most of the shipworms were obtained from mangrove wood, with the exception of *B. setacea* from unidentified found wood, and *Kuphus* spp. from both found wood and mud.

The gill metagenomes of two specimens of mud-dwelling *K. polythalamius* from Mindanao were previously sequenced (7). Here, we sequenced and analyzed gill metagenomes from a third wood-burrowing *Kuphus* sp. specimen from Mabini. Since the taxonomy of this specimen has yet to be fully resolved, we have not provided a species name for this specimen. Nonetheless, its metagenome was very similar to that of the mud dwelling *K. polythalamius* (see below). *B.*

setacea was sequenced by the JGI with low coverage. The remaining specimens from the Philippines were sequenced at Utah with highest coverage data (HiSeq), while Brazilian specimens were sequenced in Brazil and are intermediate in coverage (MiSeq) (see Table S1 for metagenome statistics). Because of their lower coverage, *B. setacea* sequences were used only in a subset of the analyses described below.

γ-Proteobacteria of Orders *Cellvibrionales* (*Teredinibacter* and allies) and *Chromatiales* (*Thiosocius* and allies) are described symbionts that live intracellularly in the gills of shipworms. We selected 23 cellulolytic and sulfur-oxidizing isolates cultivated from shipworm tissue samples. Some of the strains were previously isolated, while others originated in our recent sample collections (Table S2). For comparison, we also included *Agarilytica rhodophyticola* 017 (26) (Ga0198945), a free-living bacterium that is closely related to shipworm strains and that has a sequenced genome. The strains were sequenced at the JGI. Six of these circular genomes were closed, while remaining assemblies had between 2-141 scaffolds (see Table S2 for strain data and accession numbers). Two of these genomes, *T. turnerae* T7901 and *T. teredinicola* 2141T, were previously described (13, 21). Examination of 16S rRNA gene sequences of the cultivated strains allowed us to identify the isolates with genomes most similar to those identified in metagenome sequences (Figures 2A and S1; Table S2). Of these, 3 are cellulolytic symbionts, while 2 consist of sulfide-oxidizing symbionts. Crucially, 11 of the strains, belong to a single species, *T. turnerae*.

Each shipworm gill sample is dominated by 1-3 major bacterial species, with a large underlying strain variation. Metagenome sequencing was used to determine which bacterial species inhabit each shipworm gill. After initial analysis, we found that most of the metagenomic DNA from bacteria could be mapped to genome sequences from individual strains in our culture collection. Read counts were used to quantify the abundance of each species within each gill sample. This enabled us to accurately quantify the major symbiont species with high confidence. For this analysis, we removed *B. setacea*, which had low read coverage, and focused on the six species with good coverage (Fig. 2B).

Each shipworm gill metagenome is dominated by one to three bacterial species, which are represented by cultivated isolates that we obtained from those same species. Three shipworm species are dominated by a single bacterial species, while three others have mixed symbiont communities. As previously reported, *Kuphus* spp. is dominated by the sulfur-oxidizer *T. teredinicola*, and *N. reynei* is dominated by the cellulolytic bacterium *T. turnerae* ((7, 25). *B. thoracities* is dominated by strain 2753L. In contrast, Brazilian shipworms *Bankia* sp. and *Teredo* sp. contain mostly *T. turnerae*, but they also harbor species closely related to *Cellvibrionales* strain 1162T. *D. mannii* has the most diverse community, containing a mixture of *T. turnerae* and *Cellvibrionales* strain 2753L, and a significant fraction of the sulfide oxidizing symbiont *Chromatiales* strain 2719K. In the metagenomes of all shipworm gills, there is also a complex mixture of minor, variable bacterial strains (~1-15% of total reads, Fig. 2B shown in gray). Thus, the shipworm gill is a relatively simple system, dominated by a few species of intracellular symbiotic bacteria that are crucial to host nutrition and survival.

Underlying this simplicity, for each bacterial species present in a sample, there are multiple strain variants. We measured the prevalence of these variants using a previously reported method in which we look at single nucleotide polymorphisms in conserved genes such as DNA gyrase B at the individual read level (Figure S2) (7). The results are very similar to our previous report of bacterial strain variation in the gill metagenomes of *K. polythalamius*. The presence of multiple, closely related strain variants adds to the complexity of secondary metabolite pathways found in each host organism, since in each animal gill there are many more biosynthetic pathways than are found in individual sequenced isolates (see below).

The symbiont mixtures are representative of the animal lifestyles. *K. polythalamius* appears to thrive entirely on sulfide oxidation (7), as required in its sediment habitat, while the other shipworms contain various cellulolytic bacteria responsible for wood degradation. *D. mannii* likely has a more complex lifestyle, since it contains the sulfur-oxidizing bacterium strain 2719K and the cellulolytic species *T. turnerae* and strain 2753L. More complete descriptions of these symbioses will be published in articles focused on individual shipworm species, while here our focus is on analysis and comparison of secondary metabolism.

Shipworm symbionts and gill metagenomes are unusually rich in complex secondary metabolite pathways. The program antiSMASH (27) was used to comb the genomes for secondary metabolite BGCs. We refined the antiSMASH output to focus on BGCs that are well characterized to encode secondary metabolites: polyketide synthases (PKSs), nonribosomal peptide synthetases (NRPSs), siderophores, terpenes, homoserine lactones, and thiopeptides. Using these criteria, we identified 168 BGCs from 23 cultivated isolates and 401 BGCs from 22 shipworm gill metagenomes (Fig. 3). Because the genomes of cultivated isolates were well assembled, we could discern and analyze entire BGCs. By contrast, the animal metagenomes contained some large but many smaller contigs (N50s shown in Table S1), in which BGCs were often fragmented.

These BGCs nearly universally originate from Order *Cellvibrionales*, with very few BGCs found in the sulfide oxidizing strains *Chromatiales*. Thus, the cellulolytic shipworm symbionts are rich sources of diverse BGCs. We found only five BGCs that were similar to previously identified clusters from outside of shipworms, based upon >70% of genes conserved in antiSMASH. The remainder appeared to be unknown or uncharacterized BGCs. This result reinforces that shipworm symbiotic bacteria are promising sources of new bioactive molecules. It further supports a previous analysis comparing genomes across domain Bacteria, which revealed that *T. turnerae* represents a notably rich, yet nearly untapped, source of new secondary metabolite genes (15).

To facilitate comparison between metagenomes, we grouped BGCs into gene cluster families (GCFs). This is a method that compares groups of genes that are all involved in a particular pathway, rather than comparing individual genes (28, 29). We set an identity threshold defined in “Methods,” leading to the identification of 122 GCFs comprising all 569 discrete BGCs in the genomes and metagenomes combined (Fig. 4 and Table S3). Given that these GCFs originated in

a small handful of bacterial species, this is a large number that reinforces the role of strain variation in generating chemical diversity in shipworm gills.

A caveat is that we ignored antiSMASH hits from poorly characterized or uncharacterized biosynthetic pathway families, so that 122 GCFs represents a very conservative estimate of the chemical diversity present in shipworm gills. By paring down the genes to be analyzed to a set of well characterized BGCs, comparison became possible, but we also potentially missed some interesting pathways. As one example, we analyzed the genome of *Chromatiales* strain 2719K and discovered a gene cluster for tabtoxin (30, 31) or related compound (Fig. 5). This cluster does not contain common PKS/NRPS elements and thus was excluded from the comparative analysis (for example, it is not a defined GCF as shown in Figures 4, 6, or 7). A key biosynthetic gene in the tabtoxin-like cluster was pseudogenous in strain 2719K, but the *D. mannii* gill metagenome contained an apparently functional pathway. Tabtoxin is an important β-lactam that is used by *Pseudomonas* in plant pathogenesis (32). Since tabtoxin prevents glutamine synthesis and leads to the accumulation of toxic ammonia in plants (33), it is tempting to speculate on how tabtoxin might impact this symbiosis, perhaps improving access to nitrogen.

Another caveat is that, although we identified a large number of GCFs from a small specimen set, even this is an underestimate. *B. setacea* metagenomes from Ocean Genome Legacy were included in our analysis for comparison, but they are vastly undersampled in comparison to the other species due to low sequence coverage, and thus we are missing most of their BGCs in the analysis.

Cultivated isolate GCFs are abundant components of the source gill metagenomes. We compared BGCs and GCFs between shipworms and bacteria. Of 401 BGCs identified in the metagenomes, 305 of them also had close relatives in cultivated isolates, indicating that ~75% of BGCs in the metagenomes are covered in our sequenced culture collection (Fig. 3). Conversely, of 168 isolate BGCs, 148 (90%) of them are found in the metagenomes. Thus, sequencing further cultivated isolates in our strain collections is likely to yield additional novel BGCs.

Only 8 GCFs are widely distributed in 10 or more isolates, and these are mostly pathways that are universal or nearly universal in *T. turnerae*, which is overrepresented in our data set (Figs. 6 and 7). By contrast to isolate genomes in which we found many GCFs that occur in only a single genome, in the metagenomes most GCFS are found in multiple specimens. Only 24 out of 107 total GCFs are found only once in metagenomes (Fig. 4). Since different shipworm species in our specimen collection contain different groups of bacteria, this result reinforces the need to sample more individual specimens across the diversity of shipworm species to optimize the discovery of bioactive secondary metabolites.

To obtain a more refined view of BGC distribution, we first used the MultiGeneBlast (28) output to construct a similarity network (Fig. 6). The network provided an easily interpretable diagram of how GCFs are distributed in bacteria. However, two notable problems were observed in this diagram. First, from experience we have found the tartrolon BGC in nearly all *T. turnerae*

strains. However, this BGC was observed in only a few of the *T. turnerae*-hosting shipworms via MultiGeneBlast. This is caused by a technical problem in assembly that we often see with large *trans*-AT pathways from complex samples. Second, because this method relies upon comparing larger contigs, we could not include the low-coverage *B. setacea* sequences from JGI in this analysis.

To remedy these problems, we used a second method that obtained GCFs from cultivated isolates and used those GCFs in tBLASTn searches of metagenome contigs (Fig. 7). This provided an orthogonal view of secondary metabolism in shipworms, revealing the presence of the tartrolon pathway, as well as other pathways that do not assemble well in metagenomes because of characteristics such as repetitive DNA sequences. It also enabled us to compare *B. setacea* with the other species. A weakness of this second method is that the closeness of relationships between BGCs is not readily discerned in terms of sequence similarity and genes present. Thus, these two methods provide different insights into BGCs in shipworm gills.

Using those data, focusing on the six shipworm species with good sequencing coverage, we could observe clear trends of widely shared GCFs, GCFs that were specific to shipworm and symbiont species, and stochastically occurring GCFs.

Widely shared GCFs. Four pathways (GCF_2, GCF_3, GFC_5, and GCF_8) were prevalent in wood-eating shipworms, regardless of sample location (Figs. 6 and 7). These GCFs were encoded in the genomes of *T. turnerae* and those of several other *Cellvibrionales* isolates (especially the pathway-rich 2753L), explaining their widespread distribution.

GCF_2 encodes a NRPS / *trans*-acyltransferase (*trans*-AT) PKS pathway, the chemical products of which are unknown. It is found in all shipworm specimens in this study and in all *T. turnerae* strains. It is also present in *Cellvibrionales* strain 2753L. This explains its presence in *B. thoracites* despite the absence of *T. turnerae* in this species. GCF_2 is synonymous with "region 3" described in the annotation of the *T. turnerae* T7901 genome (13).

The most prominently occurring pathway in shipworm gill metagenomes is GCF_3. It was identified in all gill metagenomes with cellulolytic symbionts, including the metagenomes of specimen *B. setacea* BSG2. It occurs in all *T. turnerae* strains, as well as in *Cellvibrionales* strains 2753L and Bs08. It was first annotated as "region 1" in the *T. turnerae* T7901 genome and encodes an elaborate hybrid *trans*-AT PKS-NRPS pathway (13). Unlike all other GCFs identified in shipworm metagenomes and isolates, GCF_3 could be subdivided into at least three discrete categories, each of which included different gene content (Fig. 8). The first category, identified in *T. turnerae* T7901, encodes a PKS and a single NRPS, in addition to several potential modifying enzymes. Strain Bs08 contains a similar pathway, except with an additional two NRPS genes that presumably lead to additional amino acids on the resulting product. *Cellvibrionales* 2753L encoded the third pathway type, which was similar to that found in T7901 except with different flanking gene content. The presence of a single GCF that encodes similar but non-identical products suggests a dynamic pathway evolution within shipworms.

GCF_5 encodes a combination of terpene cyclase and predicted arylpolyene biosynthetic genes, which were unrecognized in the initial BGC screening of *T. turnerae* T7901 genome, since the arylpolyene pathways are recent discoveries. The GCF_5 biosynthetic product is unknown, although the cyclase and surrounding regions have all of the genes necessary to make and export hopanoids. In addition to occurring in all *T. turnerae* strains, GCF_5 is present in *Cellvibrionales* strains 1120W and 2753L. The pathway was detected in all wood-eating specimens except *Teredo* sp. TBF07 (Fig. 8).

GCF_8 is exemplified by the previously described turnerbactin BGC, from *T. turnerae* T7901. Turnerbactin is a catecholate siderophore, crucial to iron acquisition in *T. turnerae* (16). The BGC for turnerbactin was previously identified and described as "region 7" in the *T. turnerae* T7901 genome. GCF_8 is present in all *T. turnerae* genomes sequenced here. Other *Cellvibrionales* strains, including 2753L from *B. thoracites* and Bs08 from *B. setacea (*neither of which contains *T. turnerae)*, also encode turnerbactin-like siderophore synthesis. One specimen of *B. thoracites* contained GCF_8. Beyond bacterial iron acquisition, siderophores are also important in strain competition and potentially in host animal physiology (34, 35), possibly explaining the widespread distribution of GCF_8. From the clustering pattern in Figure 6, it is likely that GCF_8 comprises at least three different, but related types of gene clusters. Thus, GCF_8 likely represents catecholate siderophores, but not necessarily turnerbactin.

Important GCFs widely distributed in *T. turnerae*-containing shipworms. In addition to the four GCFs described above that have a wide distribution, GCFs 1, 4, and 11 were found in all *T. turnerae*-containing shipworms. GCF_1 is a trans-AT PKS-NRPS pathway that appears to be split into two clusters in some shipworm isolates, including *T. turnerae* T7901, in which it was previously annotated as "region 4" and "region 5". GCF_4 is the previously described "region 8" PKS-NRPS from *T. turnerae* T7901. Most notably, GCF_11 encodes tartrolon biosynthesis (17). Tartrolon is an antibiotic and potent antiparasitic agent isolated from culture broths of *T. turnerae* T7901 (17, 36, 37). It has also been identified in the cecum of the shipworm. It was proposed that the bacteria synthesize tartrolon in the gill, and it is transferred to the cecum where it may play a role in keeping the digestive tract free of bacteria (17).

GCFs specific to shipworms containing strain 2753L-like symbonts. *D. mannii* and *B. thoracites* contain abundant 2753L-like strains. Like *T. turnerae* T7901, the 2753L isolate genome encodes GCFs 2, 3, and 5. However, 2753L contains several GCFs not found in *T. turnerae*, including GCFs 6, 10, 12, 13, 14, 16, 30, and 31 (listed in order of their relative frequency of occurrence in samples). All of these GCFs are also found in *D. mannii* and *B. thoracites* gill metagenomes. These are PKS and NRPS clusters that lack close relatives according to antiSMASH annotation and thus have a potential to synthesize novel secondary metabolite classes.

GCFs specific to strain 1162T-like symbionts in shipworms from Brazil. All three Brazilian shipworm species contain symbiont genomes similar to our cultivated isolate, 1162T. This is likely due to the species sampled and not to geographical variation, since we obtained 1162T from a Philippines specimen of *Lyrodus* sp. In *Bankia* sp. and *Teredo* sp., where strain 1162T-like symbionts are a major component, there are several unique GCFs that originate in the strain

2753L-like symbionts (Fig. S4B). Because the strain 2753L-like symbiont from Brazil metagenomes is relatively distantly related to the cultivated isolate and has little overlap in terms of BGCs, we could not identify most of those full-length BGCs in the cultivated isolates. Thus, if these trends hold up through further sampling, strain 1162T-like symbionts may exhibit the high level of chemical diversity similar to what is found in strain 2753L-like genomes. In addition, when comparing the metagenomes of *Bankia* sp. and *Teredo* sp., it is clear that the 1162T-like symbionts are not identical in these two shipworm species, and that they may likely harbor different GCFs.

GCFs specific to sulfur oxidizing shipworm symbionts. *Kuphus* and its symbionts contained relatively few BGCs, but strikingly two NRPS-containing GCFs have been universally found in the shipworm associated sulfide-oxidizing species *T. teredinicola*. One of these, GCF_17, is shown in Fig. 7, where it is found in the symbiont metagenomes of *Kuphus* and *D. mannii*. It is clear that the cellulolytic symbionts are much richer in BGCs, and in addition the BGCs vary much more extensively.

Stochastic pathway occurrence in shipworm gills. Many pathways in both genomes and metagenomes were found only once or occurred relatively rarely, so that trends could not be discerned. In Fig. 7, only 18 GCFs found in more than one species of shipworm are displayed. The remaining 114 GCFs occur rarely or only once. This indicates that most biosynthetic pathways occur stochastically. This trend is reinforced in Fig. 6, where most GCFs in the diagram occur only once (single, unlinked spots). While several biosynthetic pathways are conserved and thus likely have an important conserved role in the symbiosis, most are not conserved. Further sampling of shipworm specimens, species, and cultivated isolates will yield many further, unanticipated BGCs.

Variability in conserved shipworm GCFs increases potential compound diversity. Even among conserved GCFs, there is variability in some of the pathways. This can be observed in the BGC network, where the relationships between clusters can be graphically observed. (Fig. 6). For some of these pathways, we could observe differences in gene content consistent with different chemicals that would be produced. Examples include the universal GCF_3 and siderophore pathway GCF_8 described above.

Discussion. Marine invertebrates often use small molecules in chemical defense (38). The suite of defensive chemicals in an individual animal is usually fairly simple. The production of potently bioactive compounds in large abundance (>0.1% of animal dry weight) has enabled the discovery of tens of thousands of compounds, some of which are clinically used drugs (39). Because many of the compounds resemble bacterial metabolites, chemists speculated that the true origin of compounds might be bacterial. Later work relying on genetics and metagenomics identified uncultivated bacteria as key producers of defensive metabolites (40, 41). The symbioses appear to be obligate and species-specific. To date, none of these defensive symbionts has been stably cultivated, possibly because of the long reliance on host metabolism.

On the other end of the spectrum, humans also contain many bacteria that produce potentially bioactive secondary metabolites (15, 42). Most of these have cultivated representatives. The bacteria are not obligate and are highly variable within our species. Because the resulting compounds are likely present in low abundance within a very complex microbiota, it is difficult to study the biology of secondary metabolism in humans.

Here, we describe a symbiotic system that is intermediate between these two extremes. The story began with the biology of shipworms, in which cellulolytic bacteria were long known to specifically inhabit gills and hypothesized to be the cause of an evolutionary path that leads to wood-specialization in most of the family, along with drastic morphophysiological modifications (1, 5, 43). These symbionts could be cultivated, although only recently have we been able to sample the full spectrum of major symbionts present in gills. The unexpected finding that *T. turnerae* T7901 was exceptionally rich in BGCs – proportionately denser in BGC content than *Streptomyces* spp. (13, 15) – led us to investigate shipworms as a source of new bioactive compounds. Like a considerable portion of the human microbiota, shipworm symbionts are amenable to cultivation. Here, we show that, like defensive symbioses in several marine invertebrates, the gill microbiota is relatively simple, leading to a relatively defined suite of potential metabolites. The BGCs in cultivated isolates are faithfully found as major pathways within the shipworm gill metagenome. The ability to experimentally manipulate the gill community will provide a good system to understand the role of secondary metabolism in animal symbioses.

Only four GCFs are widely conserved in cellulolytic shipworms. Two of these pathways have obvious potential roles in symbiosis. Siderophores such as turnerbactin are important in sequestering iron, but they are also widely used in bacterial competition (34, 35). Hopanoids are known to be important in symbiosis (44). Two of the pathways (GFC_2 and GCF_3) are exciting in terms of new chemistry and biology. These encode very complex biosynthetic pathways to *trans*-AT polyketides (45). Such compounds are often potently active against human disease targets. In symbioses between fungi and bacteria, *trans*-AT PKS products are highly toxic and likely defend the holobiont from predation (46). Thus, isolation of these compounds is a priority of our project.

Many pathways are specific to certain clades of shipworm symbionts, and therefore to the host animals that harbor them. These also encode very complex metabolites, most of which are polyketides, nonribosomal peptides, or hybrids of the two. There are many other such complex pathways beyond the PKS/NRPS, including tabtoxin-like pathways in specific shipworms. Most of these have yet to be characterized, with the exception of the tartrolon D/E pathway that is found in many *T. turnerae*-containing shipworms. Tartrolons are picomolar antiparasitic compounds (37), implying that they may be important in shaping the microbial environment in hosts. The specificity of BGCs to host types suggest that these pathways are important to the biology of symbiosis.

Finally, there is a large degree of stochastic occurrence of biosynthetic pathways. Chemical defensive theory suggests that differentiation or diversity of secondary metabolites increases

fitness (47). This may be especially important in highly biodiverse environments such as those found in mangrove habitats. This experimental system will enable us to examine the effects of these variable BGCs, as well as the common/universal BGCs, on the biology of the shipworm hosts.

In summary, here we show that the biosynthetic richness of cultivated shipworm symbionts is also found in the gills of the host animals. These results reveal potentially important chemical interactions that would affect a variety of marine ecosystems and a novel and underexplored source of bioactive metabolites for drug discovery.

Methods

Collection and processing of biological material. Shipworm samples (Tables S1 and S2) were collected from found wood. Briefly, infested wood was collected and transported immediately to the laboratory or stored in the shade until extraction (< 1 day). Specimens were carefully extracted to avoid damage using woodworking tools. Extracted specimens were processed immediately or stored in individual containers of filtered seawater at 4 °C until processing. Specimens were checked for viability by siphon retraction in response to stimulation and observation of heartbeat, and live specimens selected. Specimens were assigned a unique code, photographed and identified. Specimens were dissected using a dissecting stereoscope. Taxonomic vouchers (valves, pallets, and siphonal tissue for sequencing host phylogenetic markers), were retained and stored in 70% ethanol. The gill was dissected, rinsed with sterile seawater, and divided for bacterial isolation and metagenomic sequencing. Once the gill was dissected it was processed immediately or flash-frozen in liquid nitrogen.

Bacterial DNA extraction and analysis. *Teredinibacter turnerae* strains (with T prefix) were isolated using the method described in Distel el al. 2002 (12), while *Bankia setacea* symbionts (with Bs prefix) were obtained using the technique indicated in O'Connor et al. 2014 (9). Sulfur-oxidizing symbionts were isolated using the protocol specified in Altamia et al. 2019 (21). For this study, additional *T. turnerae* and novel cellulolytic symbionts from Philippine specimens (with prefix PMS) were isolated (Tables S1 and S2). Briefly, dissected gill or ceca were homogenized in sterile 75% natural seawater buffered with 20 mM HEPES, pH 8.0 using a Dounce homogenizer. Tissue homogenates were either streaked on shipworm basal medium cellulose (5) plates (1.0% Bacto Agar) or stabbed into soft agar (0.2% Bacto Agar) tubes and incubated at 25 °C until cellulolytic clearings developed. Cellulolytic bacterial colonies were subjected to several rounds of restreaking to ensure clonal selection. Contents of soft agar tubes with clearings were streaked on fresh cellulose plates to obtain single colonies. Pure colonies were then grown in 6 mL SBM cellulose liquid medium in 16 × 150 mm test tubes until the desired turbidity was observed. For long-term preservation of the isolates, a turbid medium was added to 40% glycerol at 1:1 ratio and frozen at -80 °C. Bacterial cells in the remaining liquid medium were pelleted by centrifugation at 8,000 *g* and then subjected to genomic DNA isolation. The small-subunit ribosomal (SSU) 16S rRNA gene of the isolates was then PCR amplified using 27F and 1492R from the prepared genomic DNA and sequenced. Phylogenetic analyses of 16S rRNA sequences was performed using programs implemented in Geneious,

version 10.2.3. Briefly, sequences were aligned using MAFFT (version 7.388) by using the E-INS-i algorithm. The aligned sequences were trimmed manually, resulting in a final aligned dataset of 1,125 nucleotide positions. Phylogenetic analysis was performed using FastTree (version 2.1.11) using the GTR substitution model with optimized Gamma20 likelihood and rate categories per site set to 20.

Genomic DNA used for whole genome sequencing of novel isolates and select *T. turnerae* strains were prepared using CTAB/phenol/chloroform DNA extraction method detailed in https://www.pacb.com/wp-content/uploads/2015/09/DNA-extraction-chlamy-CTAB-JGI.pdf. The purity of the extracted genomic DNA was then assessed spectrophotometrically using Nanodrop and the quantity was estimated using agarose gel electrophoresis. Samples that passed the quality control steps were submitted to Joint Genome Institute – Department of Energy (JGI-DOE) for whole genome sequencing. The sequencing platform and assembly method used to generate the final isolate genome sequences used in this study are detailed on Table S1.

Metagenomic DNA extraction.
Gill tissue samples from Philippine shipworm specimens (Table S2) were flash-frozen in liquid nitrogen and stored at -80°C prior to processing. Bulk gill genomic DNA was purified by Qiagen Blood and Tissue Genomic DNA Kit using the manufacturer's suggested protocol.

Gill tissue samples from Brazil shipworm specimens were pulverized by flash-frozen in liquid nitrogen and submitted to metagenomic DNA purification by adapting a protocol previously optimized for total DNA extraction from cnidaria tissues (48, 49). Briefly, shipworms gills were carefully dissected (taking care not to get intersections with other organs), submitted to a series of five washes with 3:1 sterile seawater / distilled water for removal of external contaminants, and macerated until powdered in liquid nitrogen. Powdered tissues (~150 mg) were then transferred to 2 mL microtubes containing 1 mL of lysis buffer [2% (m/v) cetyltrimethyl ammonium bromide (Sigma Aldrich), 1.4 M NaCl, 20 mM EDTA, 100 mM Tris-HCl (pH 8.0), with freshly added 5 µg proteinase K (v/v; Invitrogen), and 1% 2-mercaptoethanol (Sigma Aldrich)] and submitted to five freeze-thawing cycles (-80 °C to 65 °C). Proteins were extracted by washing twice with phenol:chloroform:isoamyl alcohol (25:24:1) and once with chloroform. Metagenomic DNA was precipitated with isopropanol and 5 M ammonium acetate, washed with 70% ethanol, and eluted in TE buffer (10 mM Tris-HCl, 1 mM EDTA). Metagenomic libraries were prepared using the Nextera XT DNA Sample Preparation Kit (Illumina) and sequenced with 600-cycle (300 bp paired-end runs) MiSeq Reagent Kits v3 chemistry (Illumina) at the MiSeq Desktop Sequencer.

Metagenome sequencing and assembly. *Bankia setacea* metagenomes were obtained from the JGI database (for accession numbers, see Table S1). *Kuphus polythalamius* gill metagenomes (KP2132G and KP2133G) were obtained from a previous study (7). Metagenomes from *Kuphus* sp. specimen KP3700G and *Dicyathifer mannii* and *Bactronophorus thoracites* specimens were sequenced using an Illumina HiSeq 2000 sequencer with ~350 bp inserts and 125 bp paired-end runs at the Huntsman Cancer Institute's High Throughput Genomics Center at the University of

Utah. Illumina fastq reads were trimmed using Sickle (50) with the parameters (pe sanger -q 30 –l 125). The trimmed FASTQ files were converted to FASTA files and merged using the Perl script 'fq2fq' in IBDA_ud package (51). Merged FASTA files were assembled using IDBA_ud with standard parameters in the Center for High Performance Computing at the University of Utah. For metagenome samples from Brazil, all *Neroterdo reynei* gill metagenomic samples previously analyzed were re-sequenced here to coverage depth (25). *Teredo* sp. and *Bankia* sp. gill metagenomes were sequenced using Illumina Miseq. The raw reads were assembled using either the metaspades pipeline of SPAdes (52, 53) or IDBA-UD (51). Raw reads were merged using BBMerge (54). Non-merged reads were filtered and trimmed using FaQCs (55). Both merged and processed non-merged reads were used in assembly using the metaspades pipeline.

Identification of bacterial sequences in metagenomic data. Assembly-assisted binning was used to sort and analyze trimmed reads and assembled contigs into clusters putatively representing single genomes using MetaAnnotator (56). Each binned cluster was retrieved using Samtool (57, 58). To identify bacterial clusters, genes for each bin were identified with Prodigal (59). Protein sequences for bins with coding density >50% were searched against NCBI nr database with DIAMOND (60). Bins with 60% of the genes hitting bacterial subject in the nr database were considered to originate from bacteria. Each bacterial bin was compared to the 23 shipworm isolate genomes using gANI and AF values (61). With a cut-off of AF >0.6 and gANI >0.95, the bacterial bins from each metagenome were mapped to cultivated bacterial genomes. Bins that mapped to a single bacterial genome were combined into a mega-bin. Reads mapping to each mega-bin were retrieved using MetaAnnotator.

For metagenome samples from Brazil, structural and functional annotations were carried out using DFAST (62), including only contigs with length ≥500 bp. All metagenomes were binned using Autometa (63). First, each contig's taxonomic identity was predicted using make_taxonomy_table.py, including only contigs ≥1000 bp. Predicted bacterial and archaeal contigs were binned (with recruitment via supervised machine learning) using run_autometa.py.

Building BGC similarity networks. BGCs were predicted from the bacterial contigs of each metagenome and from cultivated bacterial genomes using antiSMASH 4.0 (27). From the predictions, only BGCs for PKSs, NRPSs, siderophores, terpenes, homoserine lactones, and thiopeptides (as well as combinations of these biosynthetic enzyme families) were included in succeeding analyses. An all-versus-all comparison of these BGCs was performed using MultiGeneBlast (28) following the protocol previously reported (64). Bidirectional MultiGeneBlast BGC-to-BGC hits were considered to be reliable. In metagenome data, some truncated BGCs only showed single-directional correlation to a full length BGC. Those single-directional hits were refined as follows: protein translations of all coding sequences from the BGCs were compared in an all-versus-all fashion using blastp search. Only protein hits that had at least 60% identity and at least 80% coverage to both query and subject were considered as valid hits. A single-directional MultiGeneBlast BGC-to-BGC hit was retained if there were at least *n*-2 number of proteins (*n* is the number of proteins in the truncated BGC) passing the

blastp refining. The remaining MultiGeneBlast hits were used to construct a network in Cytoscape (65). Finally, each BGC cluster (GCF) that has relative low number of bidirectional correlations were manually checked by examining the MultiGeneBlast alignment.

Occurrence of GCFs in metagenomes. Based on the GCFs identified in previous step, the core biosynthetic proteins from each GCF were extracted and queried (NCBI tblastn) against each metagenome assembly. A threshold of query coverage of >50% and identity > 90% was applied to remove the nonspecific hits, and the remining hits, in combination with the MultiGeneBlast hits, were used to make the matrix of GCFs occurrence in metagenomes.

Acknowledgments. All collections followed Nagoya Protocol requirements; Brazilian sampling were performed under SISBIO license number 48388, and genetic resources accessed under the authorization of the Brazilian National System for the Management of Genetic Heritage and Associated Traditional Knowledge (SisGen permit number A2F0DA0). We thank the Genomics and Bioinformatics Center of Drug Research and Development Center of Federal University of Ceara for technical support.
The work was completed under supervision of the Department of Agriculture-Bureau of Fisheries and Aquatic Resources, Philippines (DA-BFAR) in compliance with all required legal instruments and regulatory issuances covering the conduct of the research. All Philippine specimens were collected under Gratuitous Permit numbers FBP-0036-10, GP-0054-11, GP-0064-12, GP-0107-15, and GP-0140-17. We thank the governments and municipalities of the Philippines and Brazil for access and help.
This work was also supported by the National Counsel of Technological and Scientific Development (CNPq) (http://cnpq.br) and by the Coordination for the Improvement of Higher Education Personnel (CAPES) (http://www.capes.gov.br) under the grant numbers 473030/2013-6 and 400764/2014-8 to AETS
Research reported in this publication was supported by the Fogarty International Center of the National Institutes of Health under Award Number U19TW008163. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The work was supported in part by US NOAA OER award #NA190AR0110303

References

1. Distel DL*, et al.* (2011) Molecular phylogeny of Pholadoidea Lamarck, 1809 supports a single origin for xylotrophy (wood feeding) and xylotrophic bacterial endosymbiosis in Bivalvia. *Mol Phylogenet Evol* 61(2):245-254.
2. Turner RD (1966) *A survey and illustrated catalogue of the Teredinidae (Mollusca: Bivalvia)* (Harvard University Press, Cambridge).
3. Distel DL, Beaudoin DJ, & Morrill W (2002) Coexistence of multiple proteobacterial endosymbionts in the gills of the wood-boring Bivalve Lyrodus pedicellatus (Bivalvia: Teredinidae). *Appl Environ Microbiol* 68(12):6292-6299.

4. Luyten YA, Thompson JR, Morrill W, Polz MF, & Distel DL (2006) Extensive variation in intracellular symbiont community composition among members of a single population of the wood-boring bivalve Lyrodus pedicellatus (Bivalvia: Teredinidae). *Appl Environ Microbiol* 72(1):412-417.
5. Waterbury JB, Calloway CB, & Turner RD (1983) A cellulolytic nitrogen-fixing bacterium cultured from the gland of deshayes in shipworms (bivalvia: teredinidae). *Science* 221(4618):1401-1403.
6. Ekborg NA, Morrill W, Burgoyne AM, Li L, & Distel DL (2007) CelAB, a multifunctional cellulase encoded by Teredinibacter turnerae T7902T, a culturable symbiont isolated from the wood-boring marine bivalve Lyrodus pedicellatus. *Appl Environ Microbiol* 73(23):7785-7788.
7. Distel DL*, et al.* (2017) Discovery of chemoautotrophic symbiosis in the giant shipworm Kuphus polythalamia (Bivalvia: Teredinidae) extends wooden-steps theory. *Proc Natl Acad Sci U S A* 114(18):E3652-E3658.
8. Betcher MA*, et al.* (2012) Microbial distribution and abundance in the digestive system of five shipworm species (Bivalvia: Teredinidae). *PLoS One* 7(9):e45309.
9. O'Connor RM*, et al.* (2014) Gill bacteria enable a novel digestive strategy in a wood-feeding mollusk. *Proc Natl Acad Sci U S A* 111(47):E5096-5104.
10. Lechene CP, Luyten Y, McMahon G, & Distel DL (2007) Quantitative imaging of nitrogen fixation by individual bacteria within animal cells. *Science* 317(5844):1563-1566.
11. Altamia MA*, et al.* (2014) Genetic differentiation among isolates of Teredinibacter turnerae, a widely occurring intracellular endosymbiont of shipworms. *Mol Ecol* 23(6):1418-1432.
12. Distel DL, Morrill W, MacLaren-Toussaint N, Franks D, & Waterbury J (2002) Teredinibacter turnerae gen. nov., sp. nov., a dinitrogen-fixing, cellulolytic, endosymbiotic gamma-proteobacterium isolated from the gills of wood-boring molluscs (Bivalvia: Teredinidae). *Int J Syst Evol Microbiol* 52(Pt 6):2261-2269.
13. Yang JC*, et al.* (2009) The complete genome of Teredinibacter turnerae T7901: an intracellular endosymbiont of marine wood-boring bivalves (shipworms). *PLoS One* 4(7):e6085.
14. Trindade-Silva AE*, et al.* (2009) Physiological traits of the symbiotic bacterium Teredinibacter turnerae isolated from the mangrove shipworm Neoteredo reynei. *Genet Mol Biol* 32(3):572-581.
15. Cimermancic P*, et al.* (2014) Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* 158(2):412-421.
16. Han AW*, et al.* (2013) Turnerbactin, a novel triscatecholate siderophore from the shipworm endosymbiont Teredinibacter turnerae T7901. *PLoS One* 8(10):e76151.
17. Elshahawi SI*, et al.* (2013) Boronated tartrolon antibiotic produced by symbiotic cellulose-degrading bacteria in shipworm gills. *Proc Natl Acad Sci U S A* 110(4):E295-304.
18. Voight JRR (2015) Xylotrophic bivalves: aspects of their biology and the impacts of humans. *J. Molluscan Stud.* 81:175-186.
19. Lopes SGBC, Domanseschi O, de Moraes DT, Morita M, & Meserani GDLC (2000) Functional anatomy of the digestive system of Neoteredo reynei (Bartsch, 1920) and Psiloteredo healdi (Bartsch, 1931) (Bivalvia: Teredinidae). *The Evolutionary Biology of*

the Bivalvia, eds Harper EM, Taylor JD, & Crame JA (Geological Society, London), Vol 177, pp 257-271.
20. Filho CS, Tagliaro CH, & Beasley CR (2008) Seasonal abundance of the shipworm Neoteredo reynei (Bivalvia, Teredinidae) in mangrove driftwood from a northern Brazilian beach. *Iheringia. Série Zoologia* 98:17-23.
21. Altamia MA, Shipway JR, Concepcion GP, Haygood MG, & Distel DL (2019) Thiosocius teredinicola gen. nov., sp. nov., a sulfur-oxidizing chemolithoautotrophic endosymbiont cultivated from the gills of the giant shipworm, Kuphus polythalamius. *Int J Syst Evol Microbiol* 69(3):638-644.
22. Shipway JR*, et al.* (2019) A rock-boring and rock-ingesting freshwater bivalve (shipworm) from the Philippines. *Proc Biol Sci* 286(1905):20190434.
23. Shipway JR*, et al.* (2016) Zachsia zenkewitschi (Teredinidae), a Rare and Unusual Seagrass Boring Bivalve Revisited and Redescribed. *PLoS One* 11(5):e0155269.
24. Elshahawi SI (2012) Isolation and biosynthesis of bioactive natural products produced by marine symbionts. PhD (Oregon Health & Science University, Portland).
25. Brito TL*, et al.* (2018) The gill-associated microbiome is the main source of wood plant polysaccharide hydrolases and secondary metabolite gene clusters in the mangrove shipworm Neoteredo reynei. *PLoS One* 13(11):e0200437.
26. Ling SK, Xia J, Liu Y, Chen GJ, & Du ZJ (2017) Agarilytica rhodophyticola gen. nov., sp. nov., isolated from Gracilaria blodgettii. *Int J Syst Evol Microbiol* 67(10):3778-3783.
27. Blin K*, et al.* (2017) antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res* 45(W1):W36-W41.
28. Medema MH, Takano E, & Breitling R (2013) Detecting sequence homology at the gene cluster level with MultiGeneBlast. *Mol Biol Evol* 30(5):1218-1223.
29. Adamek M, Spohn M, Stegmann E, & Ziemert N (2017) Mining Bacterial Genomes for Secondary Metabolite Gene Clusters. *Methods Mol Biol* 1520:23-47.
30. Kinscherf TG & Willis DK (2005) The biosynthetic gene cluster for the beta-lactam antibiotic tabtoxin in Pseudomonas syringae. *J Antibiot (Tokyo)* 58(12):817-821.
31. Kinscherf TG, Coleman RH, Barta TM, & Willis DK (1991) Cloning and expression of the tabtoxin biosynthetic region from Pseudomonas syringae. *J Bacteriol* 173(13):4124-4132.
32. Sinden SL & Durbin RD (1968) Glutamine synthetase inhibition: possible mode of action of wildfire toxin from Pseudomonas tabaci. *Nature* 219(5152):379-380.
33. Turner JG & Debbage JM (1982) Tabtoxin-induced symptoms are associated with the accumulation of ammonia formed during photorespiration. *Physiol. Plant Pathol.* 20:223-233.
34. Graf J & Ruby EG (2000) Novel effects of a transposon insertion in the Vibrio fischeri glnD gene: defects in iron uptake and symbiotic persistence in addition to nitrogen utilization. *Mol Microbiol* 37(1):168-179.
35. Holden VI & Bachman MA (2015) Diverging roles of bacterial siderophores during infection. *Metallomics* 7(6):986-995.
36. Irschik H, Schummer D, Gerth K, Hofle G, & Reichenbach H (1995) The tartrolons, new boron-containing antibiotics from a myxobacterium, Sorangium cellulosum. *J Antibiot (Tokyo)* 48(1):26-30.

37. O'Connor R & Schmidt EW (2018) PCT WO2018106966A1.
38. Schmidt EW (2008) Trading molecules and tracking targets in symbiotic interactions. *Nat Chem Biol* 4(8):466-473.
39. Molinski TF, Dalisay DS, Lievens SL, & Saludes JP (2009) Drug development from marine natural products. *Nat Rev Drug Discov* 8(1):69-85.
40. Morita M & Schmidt EW (2018) Parallel lives of symbionts and hosts: chemical mutualism in marine animals. *Nat Prod Rep* 35(4):357-378.
41. Piel J (2009) Metabolites from symbiotic bacteria. *Nat Prod Rep* 26(3):338-362.
42. Donia MS*, et al.* (2014) A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* 158(6):1402-1414.
43. Popham JD & Dickson MR (1973) Bacterial associations in the teredo Bankia australis (Lamellibranchia: Mollusca). *Mar. Biol.* 19:338-340.
44. Kulkarni G*, et al.* (2015) Specific hopanoid classes differentially affect free-living and symbiotic states of Bradyrhizobium diazoefficiens. *MBio* 6(5):e01251-01215.
45. Nguyen T*, et al.* (2008) Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat Biotechnol* 26(2):225-233.
46. Partida-Martinez LP & Hertweck C (2005) Pathogenic fungus harbours endosymbiotic bacteria for toxin production. *Nature* 437(7060):884-888.
47. Kursar TA*, et al.* (2009) The evolution of antiherbivore defenses and their contribution to species coexistence in the tropical tree genus Inga. *Proc Natl Acad Sci U S A* 106(43):18073-18078.
48. Costa-Lotufo LV*, et al.* (2018) Chemical profiling of two congeneric sea mat corals along the Brazilian coast: adaptive and functional patterns. *Chem Commun (Camb)* 54(16):1952-1955.
49. Garcia GD*, et al.* (2013) Metagenomic analysis of healthy and white plague-affected Mussismilia braziliensis corals. *Microb Ecol* 65(4):1076-1086.
50. Joshi NA & Fass JN (2011) Sickle: A sliding-window, adaptive, quality-based trimming tool for FASTQ files (Version 1.33)
51. Peng Y, Leung Hc Fau - Yiu SM, Yiu Sm Fau - Chin FYL, & Chin FY (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28(11):1420-1428.
52. Bankevich A*, et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19(5):455-477.
53. Nurk S*, et al.* (2013) Assembling Single-Cell Genomes and Mini-Metagenomes From Chimeric MDA Products. *Journal of Computational Biology* 20(10):714-737.
54. Bushnell B, Rood J, & Singer E (2017) BBMerge - Accurate paired shotgun read merging via overlap. *PLoS One* 12(10):e0185056.
55. Lo CC & Chain PS (2014) Rapid evaluation and quality control of next generation sequencing data with FaQCs. *BMC Bioinformatics* 15:366.
56. Wang Y, Leung H, Yiu S, & Chin F (2014) MetaCluster-TA: taxonomic annotation for metagenomic data based on assembly-assisted binning. *BMC Genomics* 15 Suppl 1:S12.
57. Li H*, et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078-2079.

58. Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27(21):2987-2993.
59. Hyatt D*, et al.* (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119.
60. Buchfink B, Xie C, & Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12(1):59-60.
61. Varghese NJ*, et al.* (2015) Microbial species delineation using whole genome sequences. *Nucleic Acids Research* 43(14):6761-6771.
62. Tanizawa Y, Fujisawa T, & Nakamura Y (2018) DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics* 34(6):1037-1039.
63. Miller IJ*, et al.* (2019) Autometa: automated extraction of microbial genomes from individual shotgun metagenomes. *Nucleic Acids Res* 47(10):e57.
64. Lin Z, Kakule TB, Reilly CA, Beyhan S, & Schmidt EW (2019) Secondary Metabolites of Onygenales Fungi Exemplified by Aioliomyces pyridodomos. *J Nat Prod* 82(6):1616-1626.
65. Shannon P*, et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11):2498-2504.

Figure 1.

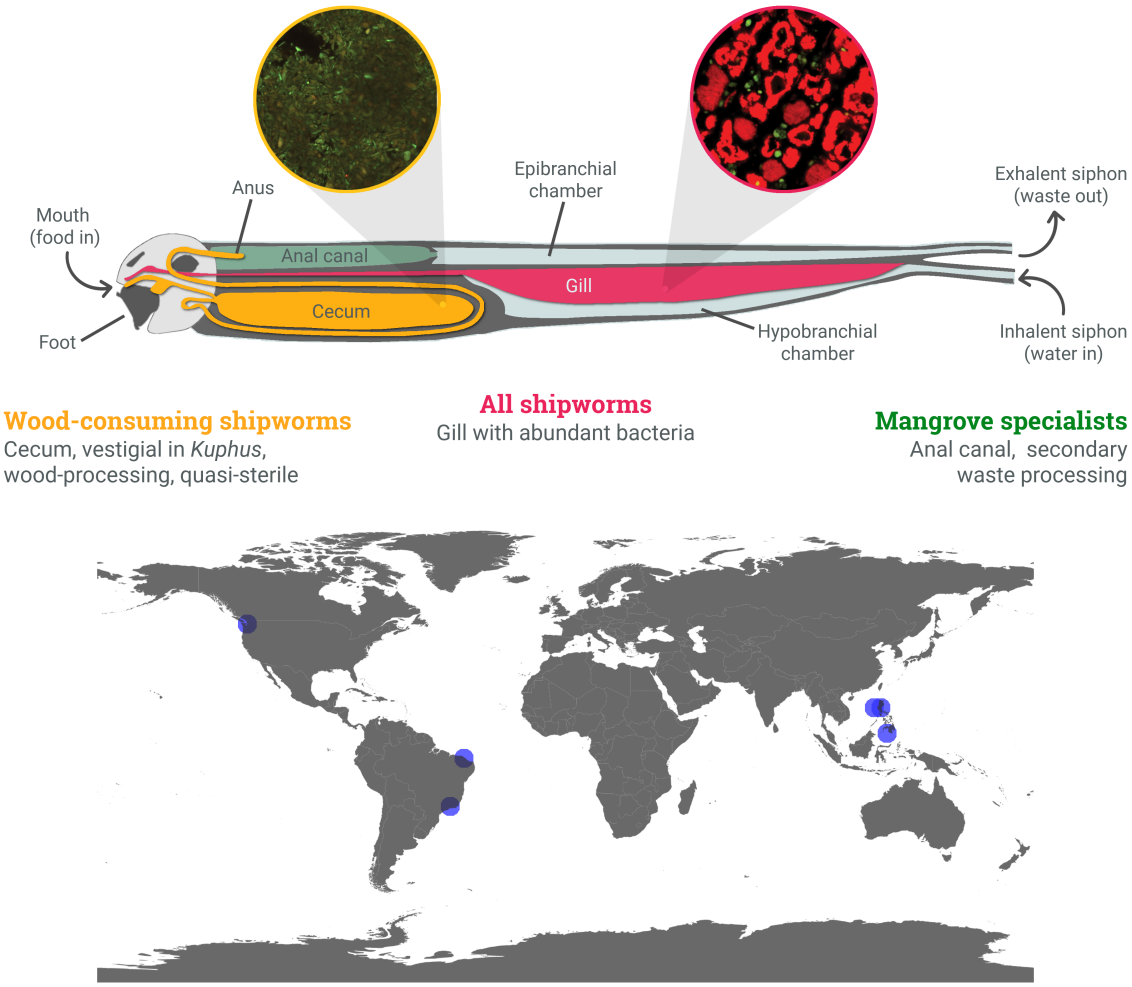


Figure 1. Top, diagram of generic shipworm anatomy. Insets are from Betcher et al., PLoS One, 2012 Figure 2, panels B and D, and are equal magnification (8). Red: signal from a fluorescent universal bacterial probe indicating large numbers of bacterial symbionts in the bacteriocytes of the gill, and paucity of bacteria in the cecum. Green is background fluorescence. Bottom, collection locations of specimens included in this study.

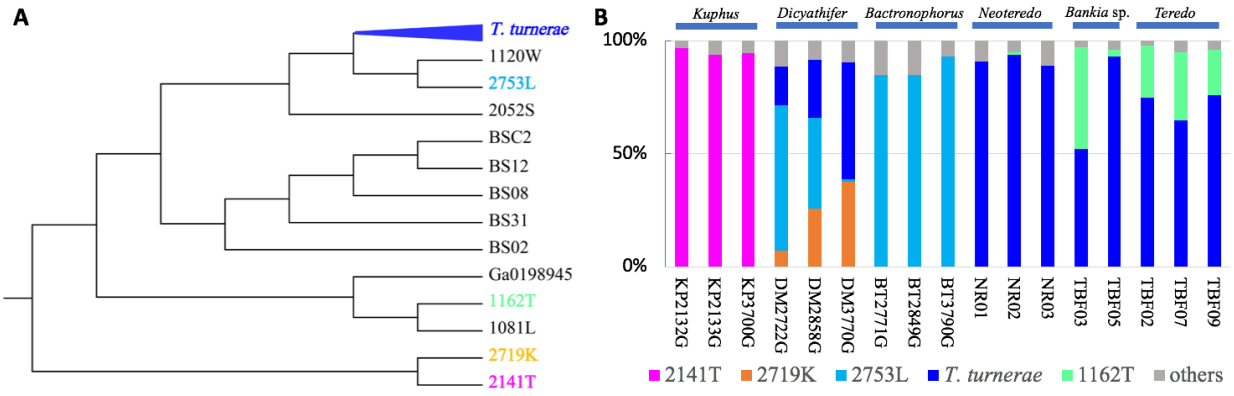


Figure 2. Cultivated bacterial isolates represent the major shipworm gill symbionts. **A)** Isolated bacteria analyzed in this study are shown in abstracted schematic of a 16S rRNA phylogenetic tree. The complete tree with accurate branch lengths and bootstrap numbers is shown in Figure S1. *T. turnerae*, within Group 1 comprised 11 sequenced strains, for other groups individual strains are shown. Each color indicates different bacteria appearing in the metagenomes in B. **B)** Species composition of shipworm gill symbiont community based on shotgun metagenome sequence analysis. The *y*-axis indicates the percent of reads originating from each bacterial species, while the *x*-axis indicates individual shipworm specimens used in the study. Colors indicate the origin of bacterial reads; gray is minor, sporadic, unidentified strains. For more details, see figure S1

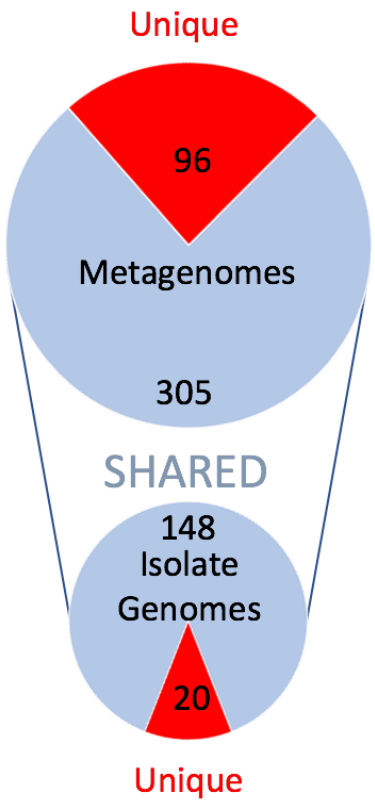


Figure 3. Most BGCs found in the metagenomes and in the bacterial isolate genomes are shared. 401 BGCs from metagenome sequences were compared to the bacterial isolate genomes, of which 305 could be found in isolates. Conversely, 148 of 168 BGCs from sequenced bacterial isolates could be found in the metagenomes. The shared numbers likely differ because the contigs assembled from the metagenome sequences were shorter on average, so that several metagenome fragments may map to a single BGC in an isolate.

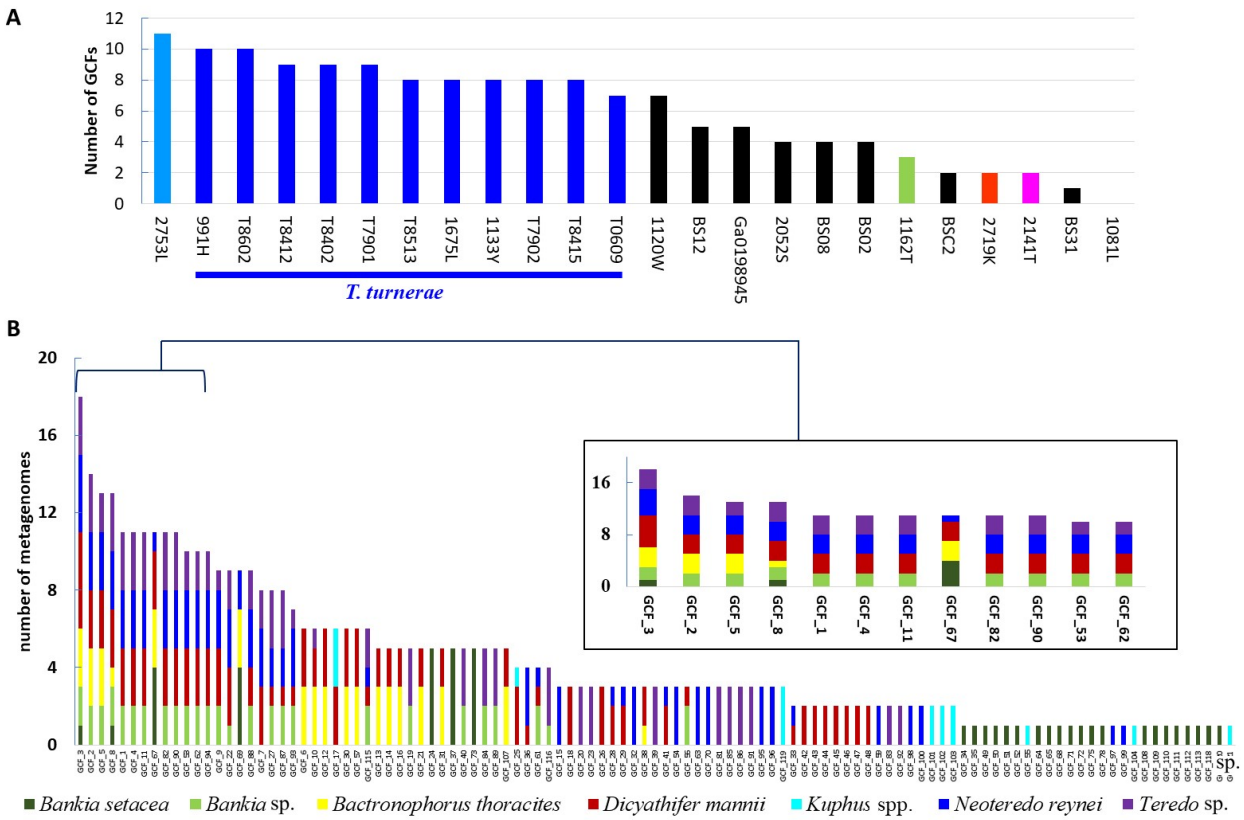


Figure 4. GCFs found in **A)** bacterial genomes and **B)** gill metagenomes. **A)** A list of GCFs found in cultivated bacterial genomes is provided in the *x*-axis, while the number of times that the GCF occurs in different sequenced strains is shown in the *y*-axis. Colors indicate bacteria from Figure 2A. Because there are 11 isolates of a *T. turnerae* in Group 1, the number of GCFs in this group (dark blue bars) are comparatively overrepresented in the diagram. **B)** GCFs (*x*-axis) found in each metagenome (*y*-axis) are shown. The inset expands a region containing the most common GCFs found in our specimens. Colors indicate shipworm host species. See Table S3 for a complete list of GCFs used in this figure.

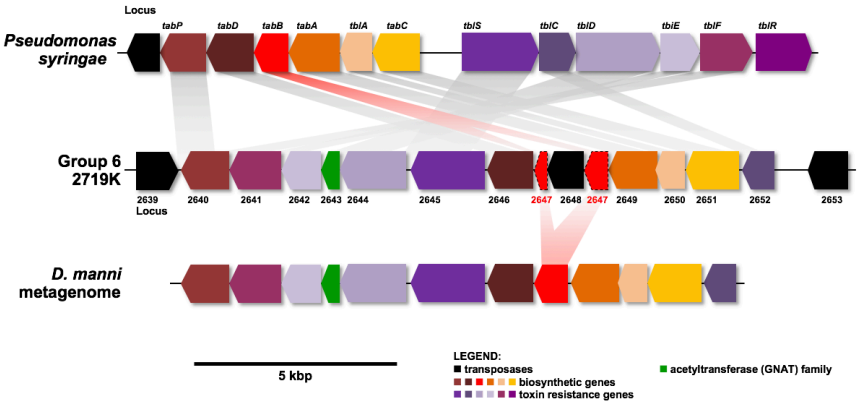


Figure 5. A possible tabtoxin pathway is found in the *D. manni* metagenome. Tabtoxin is a phytotoxin β-lactam initially discovered in *Pseudomonas* spp. (top). Strain 2719K contained a tabtoxin-like cluster that was pseudogenized (shown as an insertion in *tabB*; middle). A non-pseudogenized tabtoxin-like cluster was found in the *D. manni* metagenome gill (bottom) supporting the observation that multiple variants of each symbiont genome are represented in each metagenome.

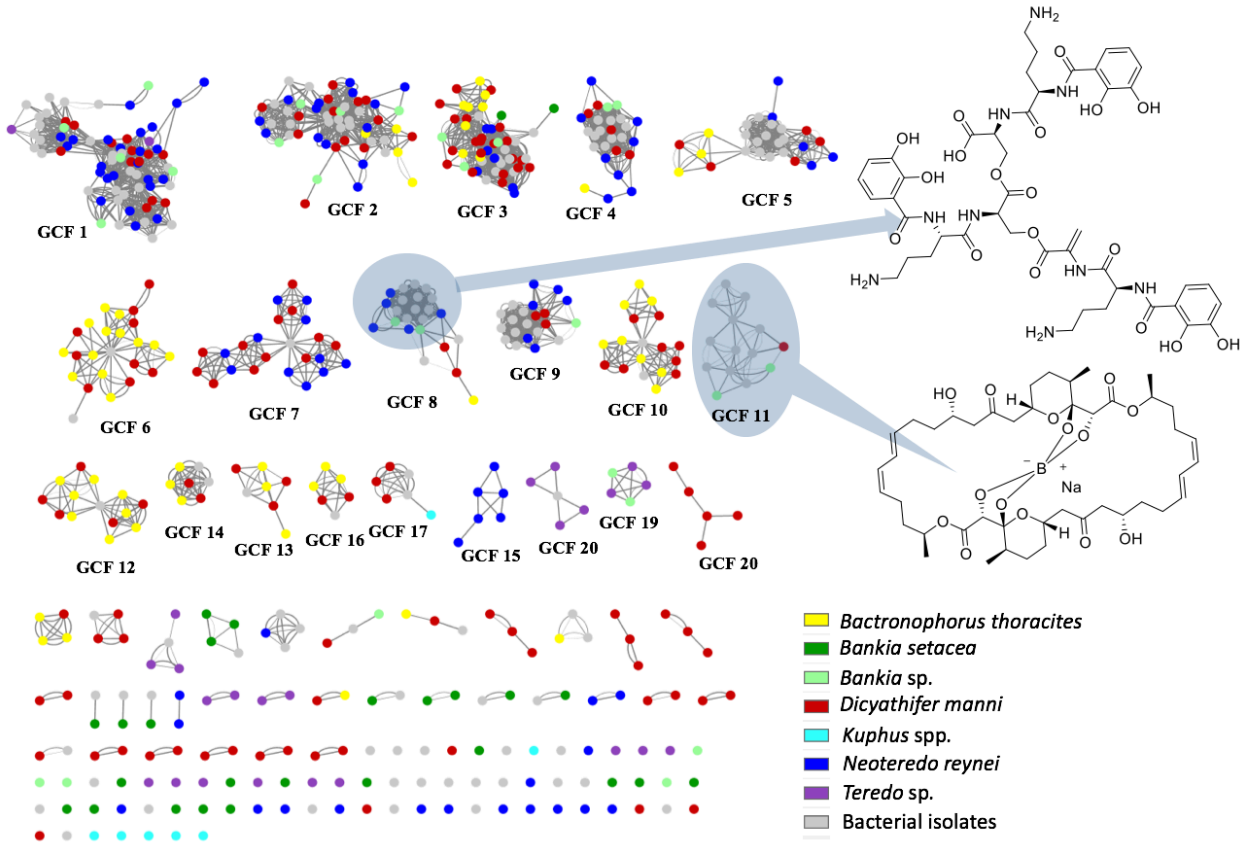


Figure 6. GCF distribution across shipworm species. Shown is a similarity network diagram, in which circles indicate individual BGCs from sequenced isolates (gray) and gill metagenomes (colors indicate species of origin; see legend). Lines indicate the MultiGeneBlast scores between identified BGCs, with thinner lines indicating a lower degree of similarity. For example, the cluster labeled "GCF_8" encodes the siderophore turnerbactin, the structure of which is shown at right. The main cluster, circled by a light blue oval, includes BGCs that are very similar to the originally described turnerbactin gene cluster. More distantly related BGCs, with fewer lines connecting them to the majority nodes in GCF_8, might represent other siderophores. GCF_11 likely all represent tartrolon D/E, a boronated polyketide shown at right. For detailed alignments of BGCs, see Fig. S3.

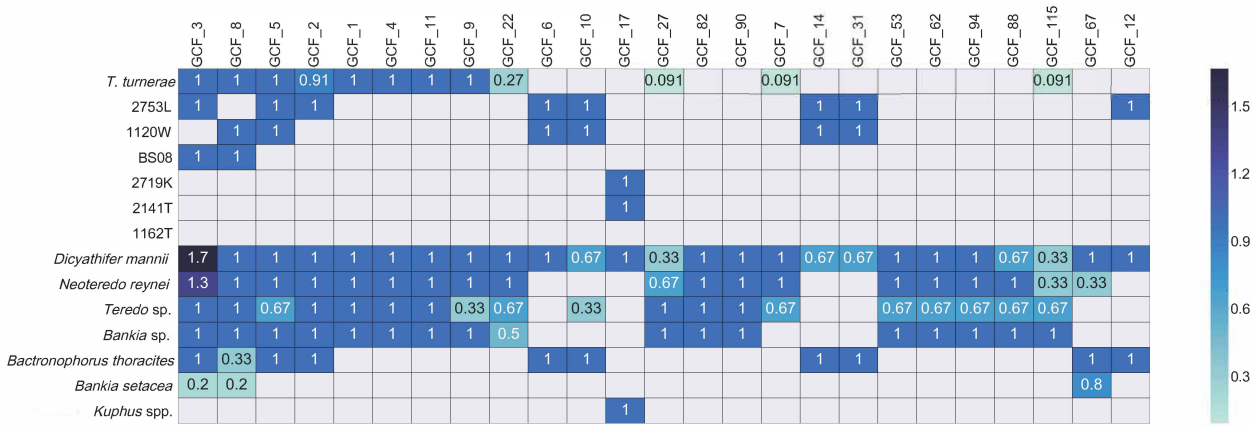


Figure 7. Pattern of occurrence of most frequently observed GCFs in isolates and metagenomes. The values in each box indicate the BGC occurrence per specimen for each GCF. For example, GCF_5 occurs in two out of three *Teredo* sp. specimens. When the number is greater than 1, as found with GCF_3, this results from having more than one of the gene cluster types present in the metagenome (see Fig. 8). When the number equals 1, the GCF is found in all specimens sampled. These data were compiled from analysis of GFCs in individual bacterial strains and samples (see Fig. S4).

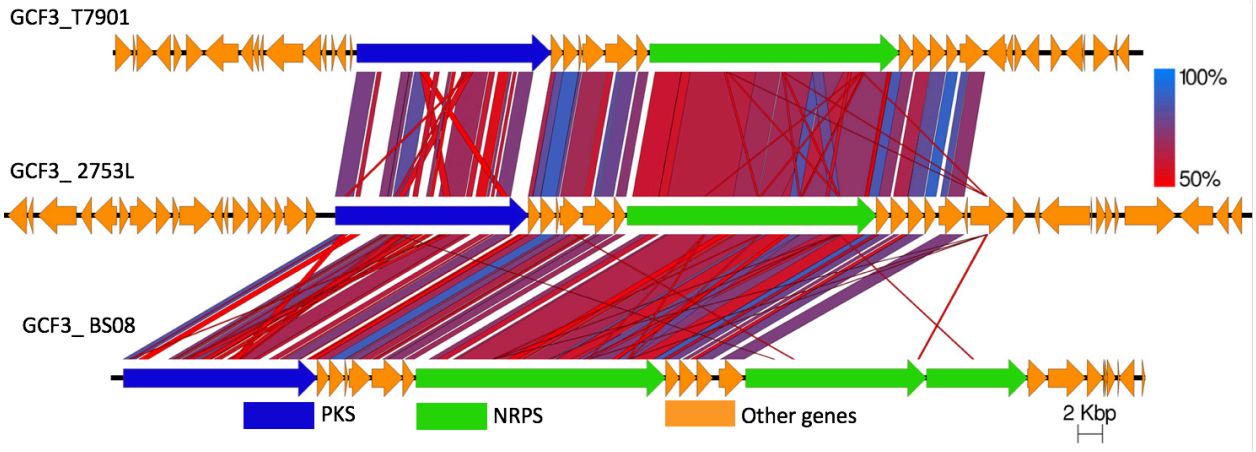


Figure 8. Three types of GCF_3 gene clusters are distributed in all cellulolytic shipworms in this study. tBLASTx was used to compare the clusters, demonstrating the presence of three closely related GCF_3 gene families found in all cellulolytic shipworm gills.

Supporting Information.

Supporting Tables: See attached docx files for Tables S1-S3.

Figure S1. Phylogeny of shipworm gill symbionts and related free-living bacteria based on approximate maximum-likelihood tree of 16S rRNA sequences. The tree was reconstructed using 1,125 nucleotide positions employing GTR substitution model in FastTree version 2.1.11 with optimized Gamma20 likelihood and rate categories per site set to 20. Support values are indicated for each node. The scale bar represents nucleotide substitution rate per site. Cultivatable shipworm symbionts and related bacteria are in boldface. The excerpted version of this tree is shown in Figure 2A.

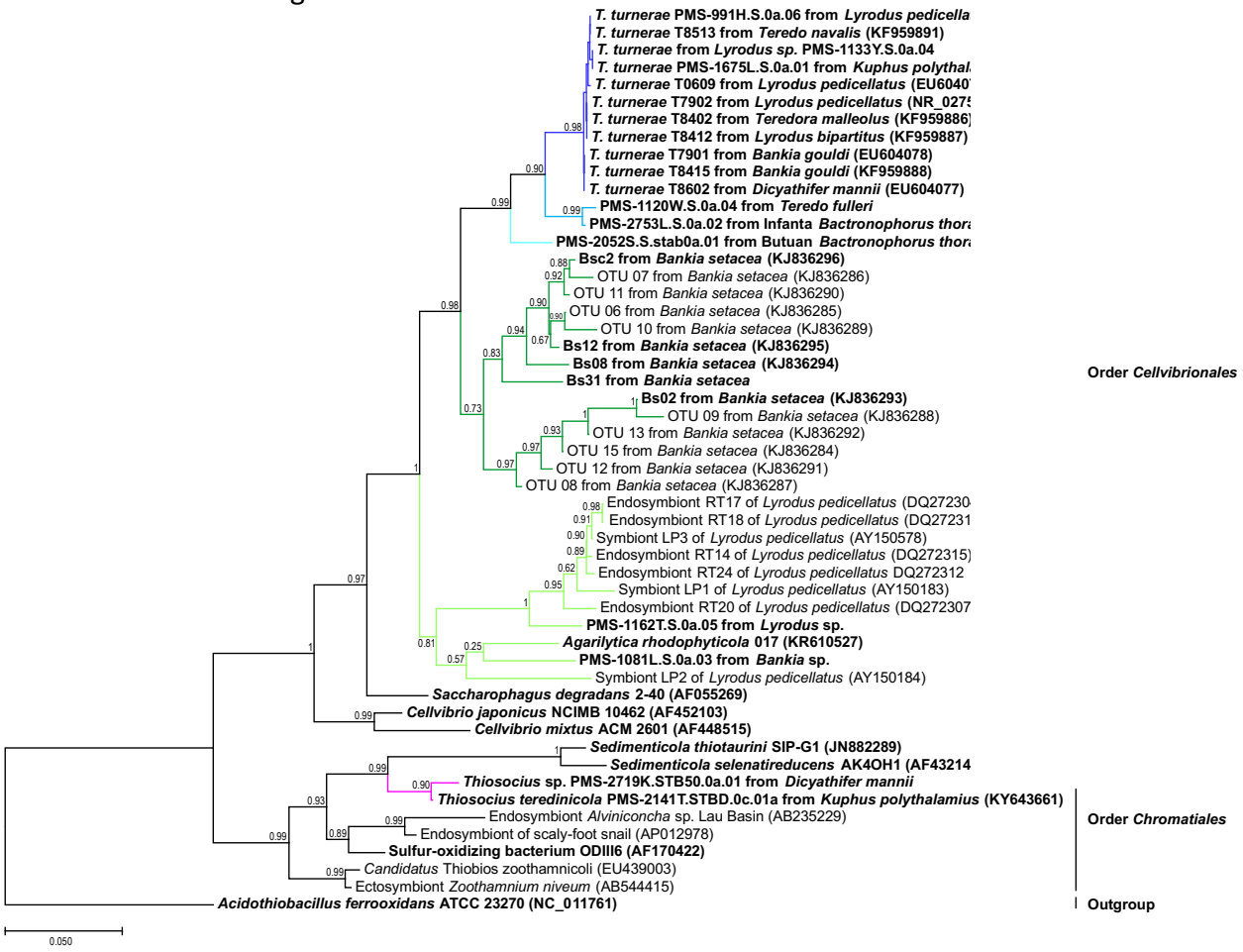


Figure S2. Strain variation in shipworm gill symbiont bacterial species. This figure was made as previously reported for *Kuphus* symbionts (7), using DNA gyrase B in 10 bp frames and examining SNP variation. Different colors indicate reads with different SNPs along the gyrase sequence. The specific example shown is from the gill of *B. thoracites* specimen 2771. The results indicate that, even though *Cellvibrionales* 2753L is the major species present in *B. thoracites* 2771, there are at minimum 2 major and 4 minor strain variants related to 2753L. The *y*-axis represents number of reads observed, while the *x*-axis indicates each 10 bp region.

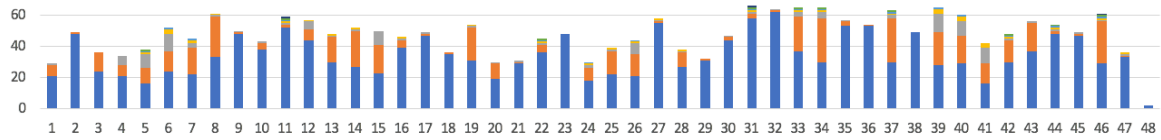
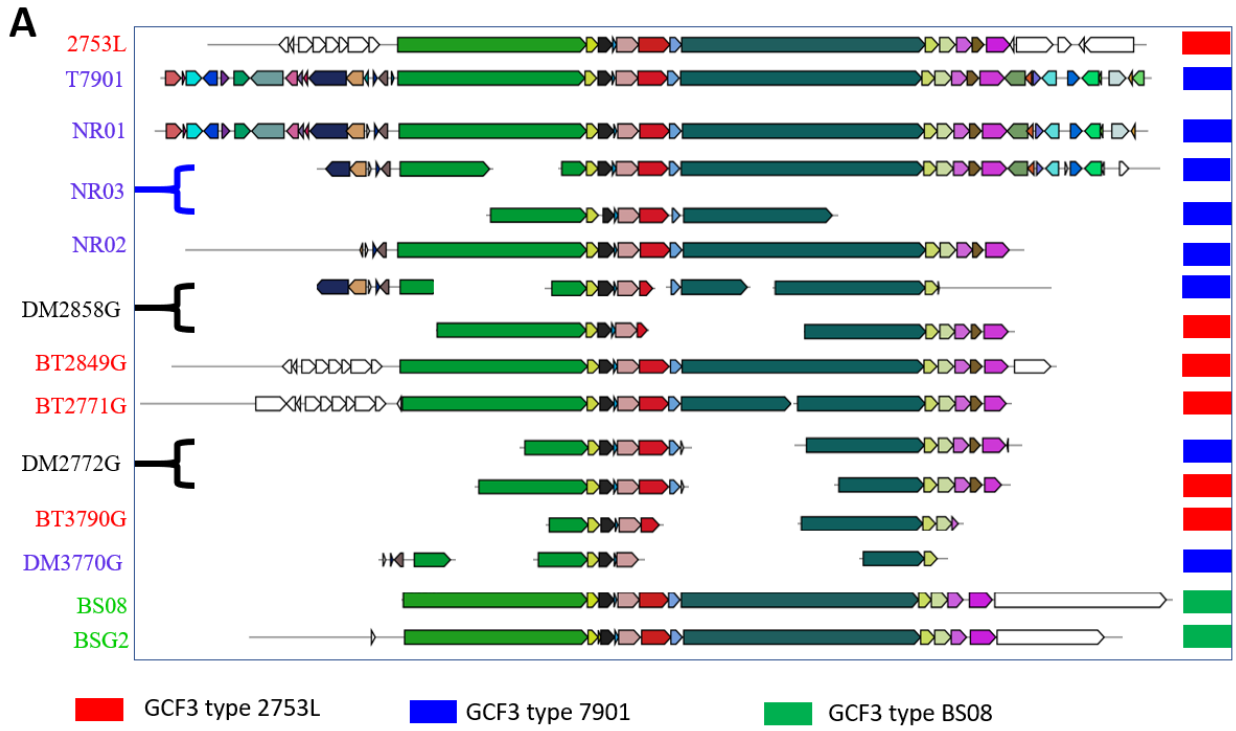
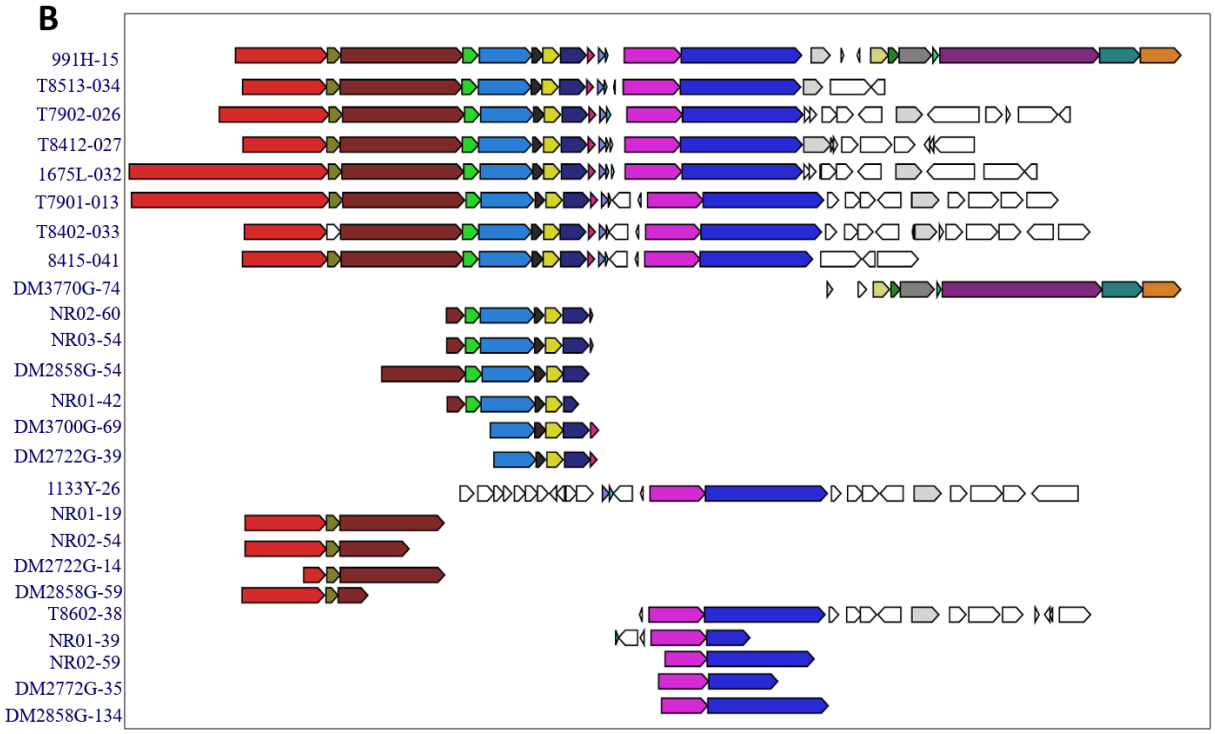
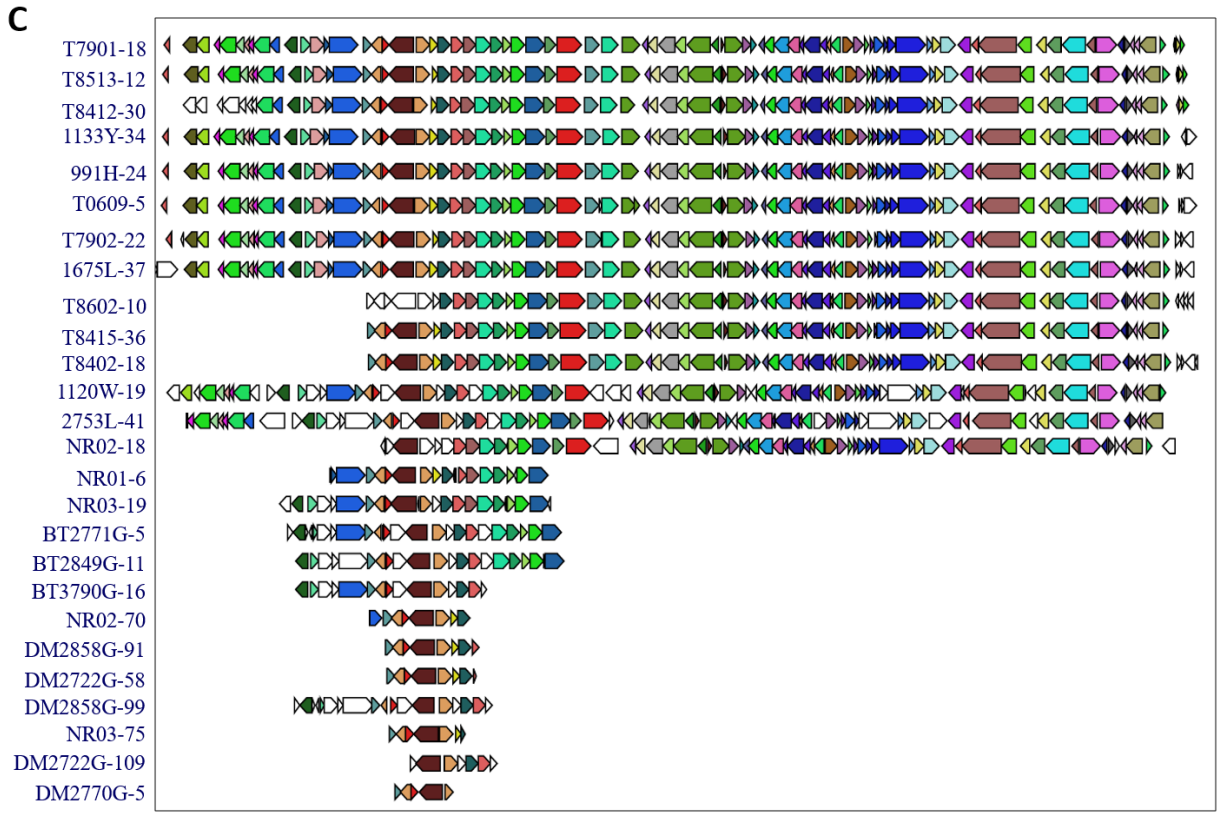


Figure S3. Representative alignments showing actual data underlying the clusters shown in Figures 3, 4, 6, and 7. **A**) representative alignment of GCF_3 from genomes and metagenomes. Three subtypes were indicated by red blue and green colors; for example, the NR03 metagenome contains two copies of blue subtype. DM2858G and DM2722G contain blue and red subtypes. **B**) alignment of GCF_2. **C**) alignment of GCF_5. **D**) alignment of GCF_8.




B
991H-15
T8513-034
T7902-026
T8412-027
1675L-032
T7901-013
T8402-033
8415-041
DM3770G-74
NR02-60
NR03-54
DM2858G-54
NR01-42
DM3700G-69
DM2722G-39
1133Y-26
NR01-19
NR02-54
DM2722G-14
DM2858G-59
T8602-38
NR01-39
NR02-59
DM2772G-35
DM2858G-134


C
T7901-18
T8513-12
T8412-30
1133Y-34
991H-24
T0609-5
T7902-22
1675L-37
T8602-10
T8415-36
T8402-18
1120W-19
2753L-41
NR02-18
NR01-6
NR03-19
BT2771G-5
BT2849G-11
BT3790G-16
NR02-70
DM2858G-91
DM2722G-58
DM2858G-99
NR03-75
DM2722G-109
DM2770G-5

D

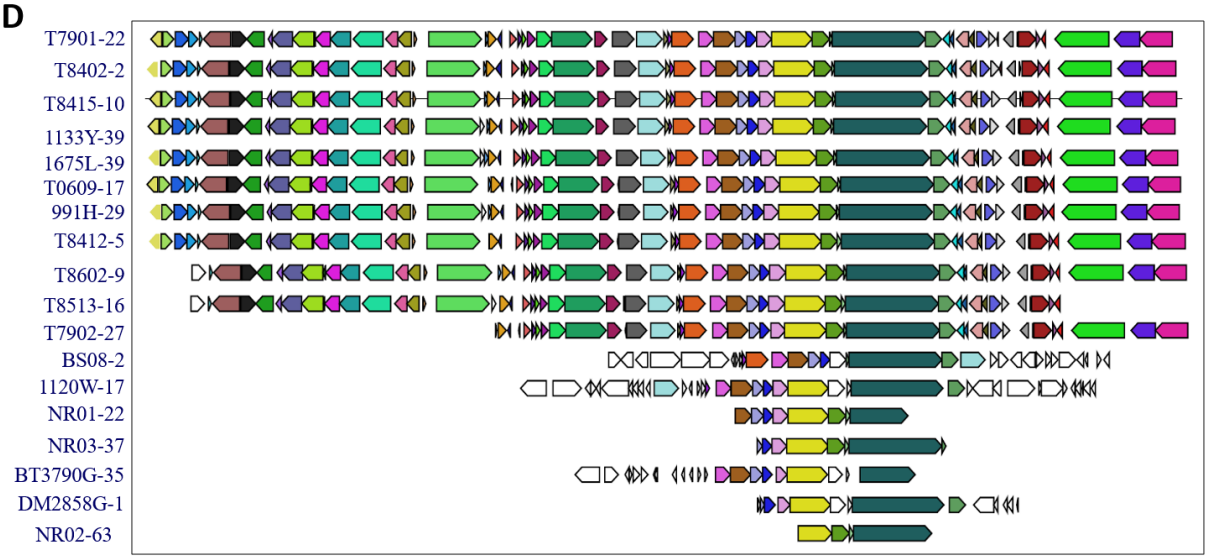
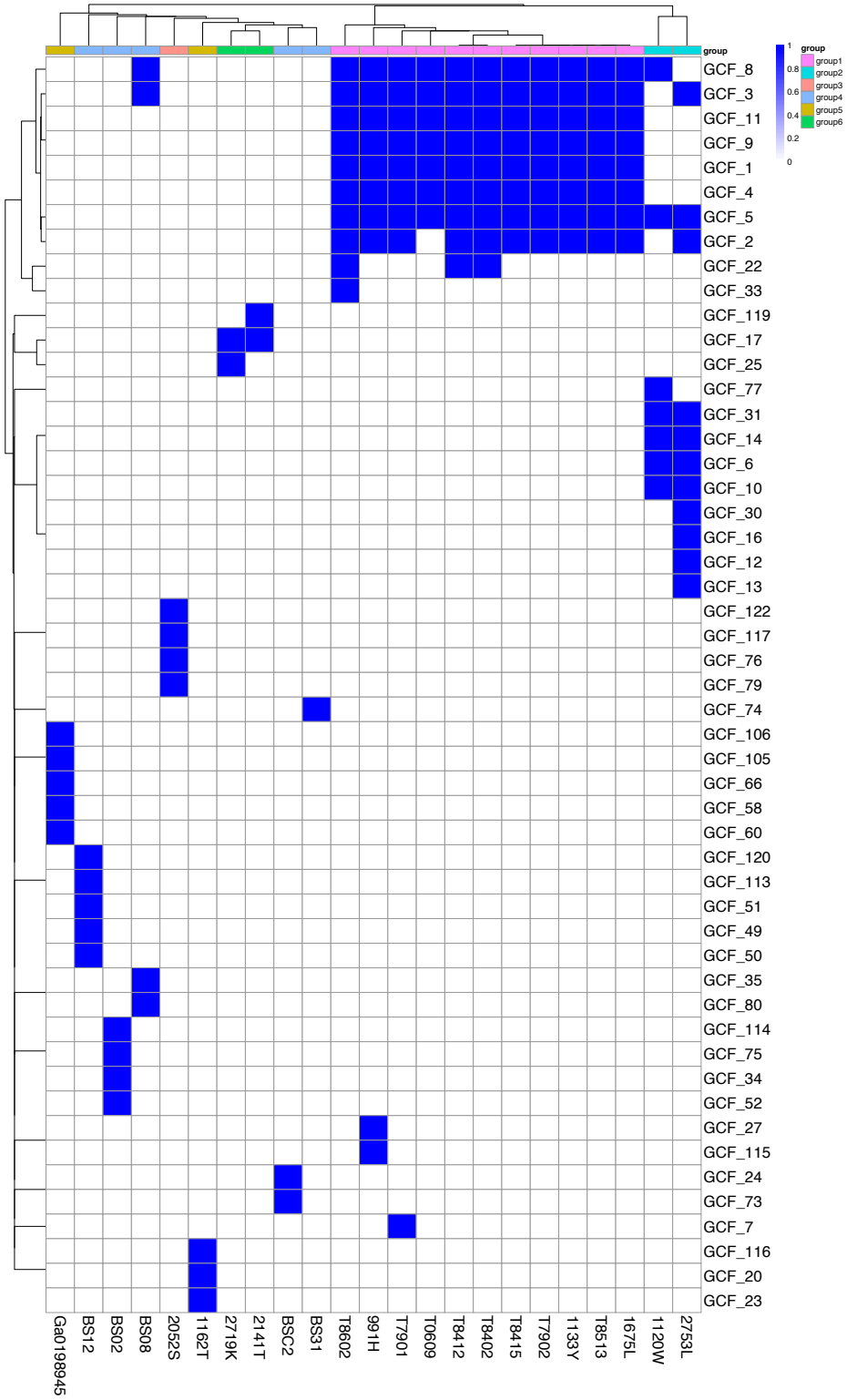

T7901-22
T8402-2
T8415-10
1133Y-39
1675L-39
T0609-17
991H-29
T8412-5
T8602-9
T8513-16
T7902-27
BS08-2
1120W-17
NR01-22
NR03-37
BT3790G-35
DM2858G-1
NR02-63

Figure S4. Occurrence of GCFs in individual samples, expanding what is shown in Fig. 7. **A)** GCFs found in bacterial strains. **B)** GCFs from individual shipworm specimens.

A



B

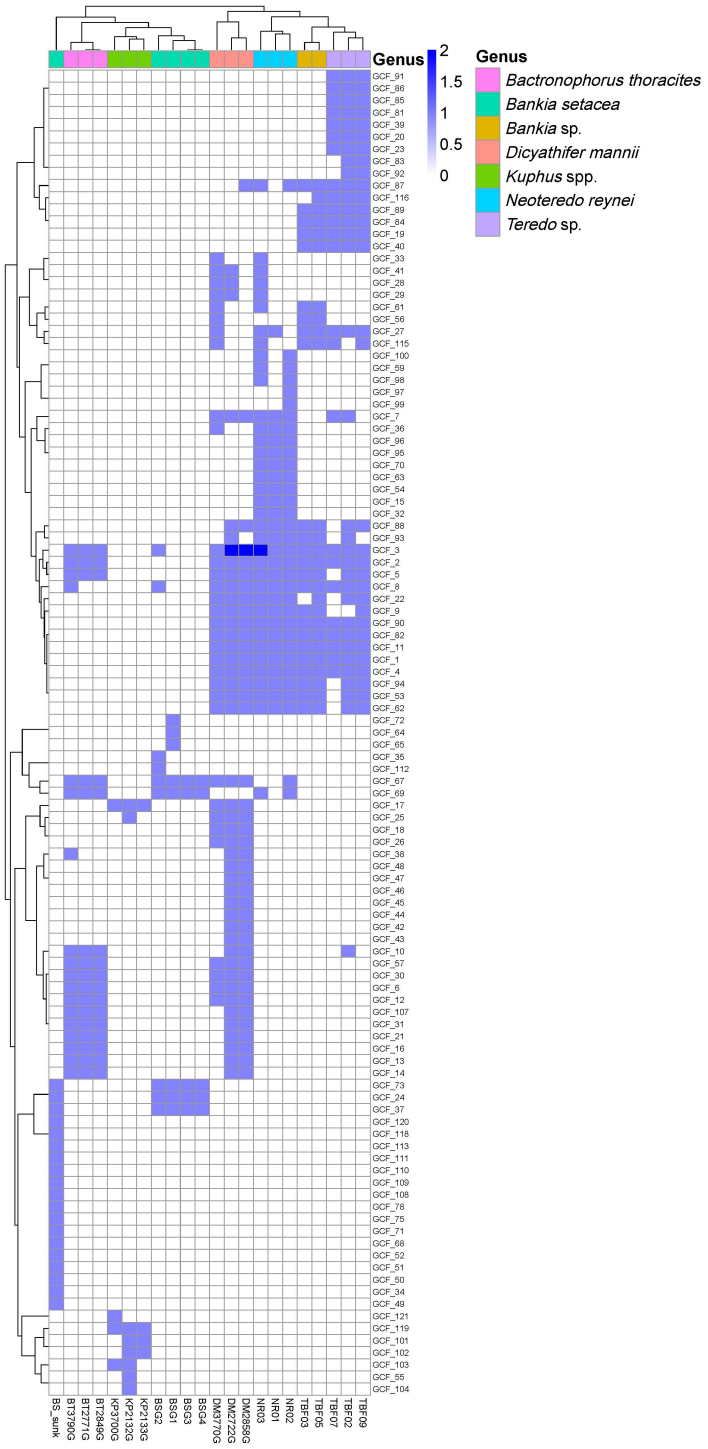

Genus
2
1.5
1
0.5
0
Genus
Bactronophorus thoracites
Bankia setacea
Bankia sp.
Dicyathifer mannii
Kuphus spp.
Neoteredo reynei
Teredo sp.
BS_sunk
BT3790G
BT2771G
BT2849G
KP3700G
KP2132G
KP2133G
BSG2
BSG1
BSG3
BSG4
DM3770G
DM2722G
DM2859G
NR03
NR01
NR02
TBF03
TBF05
TBF07
TBF02
TBF09

Table S3. List of GCFs found in this study.

GCF_1	**cf_fatty_acid-t1pks-nrps**	**GCF_62**	**terpene**
GCF_2	bacteriocin-transatpks-t1pks-nrps	GCF_63	terpene
GCF_3	cf_fatty_acid-transatpks-t1pks-nrps	GCF_64	terpene
GCF_4	t1pks-cf_saccharide-nrps	GCF_65	terpene
GCF_5	terpene-arylpolyene	GCF_66	t1pks
GCF_6	transatpks-cf_saccharide-nrps	GCF_67	t1pks
GCF_7	nrps	GCF_68	t1pks
GCF_8	cf_fatty_acid-nrps_(tunerbactin)	GCF_69	t1pks
GCF_9	t1pks	GCF_70	t1pks-PUFA
GCF_10	hserlactone-transatpks-nrps	GCF_71	t1pks-nrps
GCF_11	transatpks_(tartrolon)	GCF_72	t1pks-nrps
GCF_12	transatpks-nrps	GCF_73	t1pks-nrps
GCF_13	t1pks-nrps	GCF_74	t1pks-nrps
GCF_14	siderophore	GCF_75	t1pks-nrps
GCF_15	transatpks-otherks	GCF_76	t1pks-cf_saccharide-nrps
GCF_16	t1pks-nrps	GCF_77	t1pks-cf_saccharide-nrps
GCF_17	nrps	GCF_78	t1pks-cf_fatty_acid
GCF_18	transatpks	GCF_79	siderophore
GCF_19	t1pks	GCF_80	nrps-transatpks-otherks
GCF_20	nrps	GCF_81	nrps
GCF_21	transatpks	GCF_82	nrps
GCF_22	t1pks	GCF_83	nrps
GCF_23	siderophore	GCF_84	nrps
GCF_24	nrps	GCF_85	nrps
GCF_25	nrps	GCF_86	nrps
GCF_26	transatpks	GCF_87	nrps
GCF_27	transatpks-otherks-nrps	GCF_88	nrps
GCF_28	nrps	GCF_89	nrps
GCF_29	nrps	GCF_90	nrps
GCF_30	nrps	GCF_91	nrps
GCF_31	nrps	GCF_92	nrps

GCF_32	transatpks	GCF_93	nrps
GCF_33	transatpks-t1pks-nrps	GCF_94	nrps
GCF_34	thiopeptide-hserlactone	GCF_95	nrps
GCF_35	t1pks-cf_saccharide-nrps	GCF_96	nrps
GCF_36	t1pks-nrps	GCF_97	nrps
GCF_37	t1pks-nrps	GCF_98	nrps
GCF_38	t1pks-cf_saccharide-nrps	GCF_99	nrps
GCF_39	nrps	GCF_100	nrps
GCF_40	nrps	GCF_101	nrps
GCF_41	nrps	GCF_102	nrps
GCF_42	nrps	GCF_103	nrps
GCF_43	nrps	GCF_104	nrps
GCF_44	nrps	GCF_105	nrps
GCF_45	nrps	GCF_106	nrps
GCF_46	nrps	GCF_107	nrps
GCF_47	nrps	GCF_108	nrps
GCF_48	nrps	GCF_109	nrps
GCF_49	nrps	GCF_110	nrps
GCF_50	nrps	GCF_111	nrps
GCF_51	hserlactone-t1pks-nrps	GCF_112	nrps
GCF_52	cf_saccharide-nrps	GCF_113	nrps
GCF_53	transatpks	GCF_114	nrps
GCF_54	transatpks	GCF_115	nrps
GCF_55	transatpks	GCF_116	nrps
GCF_56	transatpks	GCF_117	hserlactone-transatpks-cf_fatty_acid
GCF_57	transatpks	GCF_118	hserlactone-nrps
GCF_58	transatpks-t1pks-nrps	GCF_119	cf_saccharide-nrps
GCF_59	transatpks-otherks	GCF_120	cf_fatty_acid-t1pks
GCF_60	transatpks-cf_saccharide	GCF_121	bacteriocin-lantipeptide
GCF_61	transatpks-cf_fatty_acid	GCF_122	arylpolyene-nrps_(butunamide)

Table S1. Shipworm gill metagenomes used in this study.

#	Gill metagenome	PMS-ICBG sample codes	Source shipworm species	Location	Coordinates	Sequencing center	Sequencing platform	Assembler	Reads, post trim	Size in bp	No. of contigs	N50	%GC	IMG Genome ID
1	DM2722G	PMS-2722P	*Dicyathifer mannii* specimen PMS-2717Y	Infanta, Quezon, Philippines	N 14.68367°, E 121.63690°	Huntsman Cancer Institute, University of Utah	Illumina HiSeq 2000	IDBA_ud	187291588	1235295176	924064	2095	34.9	
2	BT2771G	PMS-2771X	*Bactronophorus thoracites* specimen PMS-2769U	Infanta, Quezon, Philippines	N 14.68367°, E 121.63690°	Huntsman Cancer Institute, University of Utah	Illumina HiSeq 2000	IDBA_ud	177392546	1056707310	813604	2023	35.4	
3	BT2849G	PMS-2849Y	*Bactronophorus thoracites* specimen PMS-2839H	Infanta, Quezon, Philippines	N 14.68367°, E 121.63690°	Huntsman Cancer Institute, University of Utah	Illumina HiSeq 2000	IDBA_ud	193099534	1059162705	814617	2024	35.4	
4	DM2858G	PMS-2858W	*Dicyathifer mannii* specimen PMS-2823T	Infanta, Quezon, Philippines	N 14.68367°, E 121.63690°	Huntsman Cancer Institute, University of Utah	Illumina HiSeq 2000	IDBA_ud	186697500	1236681788	928980	2083	34.9	
5	DM3770G	PMS-3770U	*Dicyathifer mannii* specimen PMS-3768S	Infanta, Quezon, Philippines	N 14.68367°, E 121.63690°	Huntsman Cancer Institute, University of Utah	Illumina HiSeq 2000	IDBA_ud	297553066	1328488478	1067922	1946	34.9	
6	BT3790G	PMS-3790S	*Bactronophorus thoracites* specimen PMS-3779S	Infanta, Quezon, Philippines	N 14.68367°, E 121.63690°	Huntsman Cancer Institute, University of Utah	Illumina HiSeq 2000	IDBA_ud	309554332	1127330840	927279	1873	35.4	
10	KP3700G	PMS-3700M	*Kuphus* sp. specimen PMS-3696Y (wood-boring)	Mabini, Batangas, Philippines	N 13.75843°, E 120.92586°	Huntsman Cancer Institute, University of Utah	Illumina HiSeq 2000	IDBA_ud	82015762	734092095	358482	4300	37.6	
11	KP2132G	PMS-2246K and PMS-2249P	*Kuphus polythalamius* specimen PMS-2132W (mud-dwelling)	Kalamansig, Sultan Kudarat, Philippines	N 6.53631°, E 124.048365°	Huntsman Cancer Institute, University of Utah	Illumina HiSeq 2000	IDBA_ud	318294870	772720664	424816	4530	37.6	
12	KP2133G	PMS-2157H, PMS-2116M, and PMS-2110W	*Kuphus polythalamius* specimen PMS-2133X (mud-dwelling)	Kalamansig, Sultan Kudarat, Philippines	N 6.53631°, E 124.048365°	Huntsman Cancer Institute, University of Utah	Illumina HiSeq 2000	IDBA_ud	329174268	795400237	500141	3879	37.4	
13	BSG1	-	*Bankia setacea*	Puget Sound, Washington, USA	N 47.85072°, W 122.33843°	Joint Genome Institute - Department of Energy	Illumina HiSeq 2000	SOAPdenovo, Newbler, and Minimus2	-	563042012	761912	985	35.0	3300000111
14	BSG3	-	*Bankia setacea*	Puget Sound, Washington, USA	N 47.957498°, W 122.529373°	Joint Genome Institute - Department of Energy	Illumina HiSeq 2000	SOAPdenovo, Newbler, and Minimus2	-	620222960	648493	1550	34.9	3300000024
15	BSG2	-	*Bankia setacea*	Puget Sound, Washington, USA	N 47.957498°, W 122.529373°	Joint Genome Institute - Department of Energy	Illumina HiSeq 2000	SOAPdenovo, Newbler, and Minimus2	-	540217764	793976	860	34.8	3300000110
17	BSG4	-	*Bankia setacea*	Puget Sound, Washington, USA	N 47.85072°, W 122.33843°	Joint Genome Institute - Department of Energy	Illumina HiSeq 2000	SOAPdenovo, Newbler, and Minimus2	-	574332630	692986	1194	34.6	3300000107
19	BS_sunk	-	*Bankia setacea*	Puget Sound, Washington, USA	N 47.85072°, W 122.33843	Joint Genome Institute - Department of Energy	Illumina, 454 GS FLX Titanium	Newbler and Velvet	-	26539887	38227	1943	45.2	2070309010
20	NR01	-	*Neoteredo reynei*	Coroa grande Mangrove - Sepetiba bay, Rio de Janeiro State, BR	22.9081670° S 43.8756390° W	CEGENBIO	Illumina MiSeq	SPAdes	9224156	313630826	413893	779	37.3	
21	NR02	-	*Neoteredo reynei*	Coroa grande Mangrove - Sepetiba bay, Rio de Janeiro State, BR	22.9081670° S 43.8756390° W	CEGENBIO	Illumina MiSeq	SPAdes	18338062	416566737	468503	986	37.2	
22	NR03	-	*Neoteredo reynei*	Coroa grande Mangrove - Sepetiba	22.9081670° S 43.8756390° W	CEGENBIO	Illumina MiSeq	SPAdes	13078802	309408486	414159	769	38.2	

				bay, Rio de Janeiro State, BR										
23	TBF02	-	*Teredo* sp.	Environmental Preservation Area of Pacoti river, Ceará State, Brazil	S 3.843111, W 38.422695 (3°50'35.2"S 38°25'21.7"W)	CEGENBIO	Illumina MiSeq	SPAdes	3565711	74108472	236018	1037	40.1	
24	TBF03	-	*Bankia* sp.	Environmental Preservation Area of Pacoti river, Ceará State, Brazil	S 3.843111, W 38.422695 (3°50'35.2"S 38°25'21.7"W)	CEGENBIO	Illumina MiSeq	SPAdes	2205607	33524312	75538	1123	41.2	
25	TBF05	-	*Bankia* sp.	Environmental Preservation Area of Pacoti river, Ceará State, Brazil	S 3.843111, W 38.422695 (3°50'35.2"S 38°25'21.7"W)	CEGENBIO	Illumina MiSeq	SPAdes	3632367	107179837	230494	995	37.4	
26	TBF07	-	*Teredo* sp.	Environmental Preservation Area of Pacoti river, Ceará State, Brazil	S 3.843111, W 38.422695 (3°50'35.2"S 38°25'21.7"W)	CEGENBIO	Illumina MiSeq	SPAdes	3731031	78684542	258368	965	38.6	
27	TBF09	-	*Teredo* sp.	Environmental Preservation Area of Pacoti river, Ceará State, Brazil	S 3.843111, W 38.422695 (3°50'35.2"S 38°25'21.7"W)	CEGENBIO	Illumina MiSeq	SPAdes	4029653	108441874	340072	948	38.1	

Table S2: Shipworm symbiont genomes.

#	Code in the manuscript	Isolate name	Metabolic type	Host shipworm	Location	Coordinates	Sequencing center	Sequencing platform	Sequence assembler	Estimated genome size	No. of contigs/scaffolds	N50	%GC	IMG Genome ID
1	T7901	*T. turnerae* strain T7901	Cellulolytic	*Bankia gouldi*	Beaufort, North Carolina USA	N 34.71737°, W 76.67198°	J. Craig Venter Institute	454, Sanger	Celera Assembler and custom software	5,193,164	1 (closed circular)	Not applicable	50.89	2541046951
2	T8415	*T. turnerae* strain T8415	Cellulolytic	*Bankia gouldi*	Fort Pierce, Florida, USA	N 27.48063°, W 80.30967°	JGI-DOE	Illumina	ALLPATHS	5,158,349	50	Scaffold N/L50: 5/398.1 Kbp Contig N/L50: 6/395.4 kbp	50.78	2510917000
3	T8602	*T. turnerae* strain T8602	Cellulolytic	*Dicyathifer mannii*	Townsville, Queensland, Australia	S 19.27631°, E 147.05784°	JGI-DOE	Illumina	ALLPATHS	5,097,488	59	Scaffold N/L50: 6/291.7 kbp Contig N/L50: 2/291.7 kbp	51.03	2513237135
4	T7902	*T. turnerae* strain T7902	Cellulolytic	*Lyrodus pedicellatus*	Long Beach, California, USA	N 33.76138°, W 118.17281°	JGI-DOE	Illumina	ALLPATHS	5,387,817	72	Scaffold N/L50: 11/176.4 kbp Contig N/L50: 11/176.4 kbp	50.81	2513237099
5	T8402	*T. turnerae* strain T8402	Cellulolytic	*Teredora malleolus*	Floating wood in the Atlantic Ocean	N 38.30667°, W 69.59333°	JGI-DOE	Illumina	Velvet (1.1.04) and ALLPATHS-LG	5,166,130	27	Scaffold N/L50: 6/348.4 kbp Contig N/L50: 7/315.4 kbp	50.86	2519899652
6	T8412	*T. turnerae* strain T8412	Cellulolytic	*Lyrodus bipartitus*	Jim Island, Fort Piece, Florida, USA	N 27.476944°, W 80.311944°	JGI-DOE	Illumina	Velvet (1.1.04) and ALLPATHS-LG	5,147,360	58	Scaffold N/L50: 10/205.3 kbp Contig N/L50: 10/205.3 kbp	51.07	2519899664
7	T0609	*T. turnerae* strain T0609	Cellulolytic	*Lyrodus pedicellatus*	Long Beach, California, USA	N 33.76138°, W 118.17281°	JGI-DOE	Illumina	Velvet (1.1.04) and ALLPATHS-LG	5,069,061	49	Scaffold N/L50: 7/246.6 kbp Contig N/L50: 7/246.6 kbp	51.15	2519899663
8	991H	*T. turnerae* strain PMS-991H.S.0a.06	Cellulolytic	*Lyrodus pedicellatus* specimen PMS-988W	Panglao, Bohol, Philippines	N 9.54558°, E 123.76030°	JGI-DOE	Illumina	ALLPATHS-LG	5,279,031	13	Scaffold N/L50: 2/1.8 Mbp Contig N/L50: 3/888.4 kbp	51.07	2524614873

9	T8513	*T. turnerae* strain T8513	Cellulolytic	*Teredo navalis*	São Paulo, Brazil	S 23.81992°, W 45.40517°	JGI-DOE	Illumina	Velvet (1.1.04) and ALLPATHS-LG	5,268,281	84	Scaffold N/L50: 9/189.8 kbp Contig: 8/189.8 kbp	50.92	2523533596
10	1133Y	*T. turnerae* strain PMS-1133Y.S.0a.04	Cellulolytic	*Lyrodus* sp. specimen PMS-1128S	Panglao, Bohol, Philippines	N 9.59670°, E 123.74990°	JGI-DOE	Illumina	ALLPATHS-LG	5,134,977	6	Scaffold N/L50: 1/3.2 Mbp Contig N/L50: 4/607.0 kbp	50.85	2540341229
11	1675L	*T. turnerae* strain PMS-1675L.S.0a.01	Cellulolytic	*Kuphus polythalamius* specimen PMS-1672Y	Kalamansig, Sultan Kudarat, Philippines	N 6.53631°, E 124.04836°	JGI-DOE	PacBio	HGAP 2.1.1	5,283,781	1 (closed circular)	Not applicable	51.05	2571042908
12	2753L	PMS-27553L.S.0a.02	Cellulolytic	*Bactronophorus thoracites* specimen PMS-2749X	Infanta, Quezon, Philippines	N 14.68367°, E 121.63690°	JGI-DOE	PacBio	HGAP 2.1.1	6,056,039	2	Scaffold N/L50: 1/4.4 Mbp	47.96	2579779156
13	1120W	PMS-1120W.S.0a.04	Cellulolytic	*Teredo fulleri* specimen PMS-1114L	Panglao, Bohol, Philippines	N 9.59670°, E 123.74990°	JGI-DOE	PacBio	HGAP 2.0.1	5,699,307	1 (closed circular)	Not applicable	50.39	2558309032
14	2052S	PMS-2052S.S.stab0a.01	Cellulolytic	*Bactronophorus thoracites* specimen PMS-1959H	Butuan, Agusan del Norte, Philippines	N 8.98650°, E 125.45768°	JGI-DOE	Illumina	ALLPATHS -LG	5,635,926	3	Scaffold N/L50: 1/5.6 Mbp Contig: 3/981.6 kbp	54.68	2541046951
15	Bs12	Bs12	Cellulolytic	*Bankia setacea*	Puget Sound, Washington, USA	N 47.95749°, W 122.52937°	JGI-DOE	PacBio	HGAP 2.0.0	4,921,245	3	Contig: 1/4.7 Mbp	45.72	2545555829
16	Bs08	Bs08	Cellulolytic	*Bankia setacea*	Puget Sound, Washington, USA	N 47.95749°, W 122.52937°	JGI-DOE	Illumina	Velvet v. DEC-2010	4,814,259	90	Scaffold N/L50: 7/255.3 Mbp Contig: 14/112.2 kbp	47.18	2767802764
17	Bsc2	Bsc2	Cellulolytic	*Bankia setacea*	Puget Sound, Washington, USA	N 47.95749°, W 122.52937°	New England Biolabs	PacBio	HGAP 2.0.1	5,414,953	10	4.2 Mbp	47.31	2531839719
18	Bs31	Bs31	Cellulolytic	*Bankia setacea*	Puget Sound, Washington, USA	N 47.95749°, W 122.52937°	JGI-DOE	PacBio	Velvet 1.1.04 and ALLPATHS-LG	5,017,353	46	Scaffold N/L50: 5/341.1 kbp Contig: 8/260.1 kbp	47.60	2528768159
19	Bs02	Bs02	Cellulolytic	*Bankia setacea*	Puget Sound, Washington, USA	N 47.95749°, W 122.52937°	JGI-DOE	Illumina	Velvet v. DEC-2010	3,886,134	141	Contig: 8/176.2 kbp	47.76	2503982003
20	1162T	PMS-1162T.S.0a.05	Cellulolytic	*Lyrodus* sp. specimen PMS-1157K	Talibon, Bohol, Philippines	N 10.30748°, E 124.40168°	JGI-DOE	Illumina and PacBio	ALLPATHS-LG	4,404,964	1 (closed circular)	Not applicable	47.72	2524614822
21	1081L	PMS-1081L.S.0a.03	Agarolytic	*Bankia* sp. specimen PMS-1083P	Panglao, Bohol, Philippines	N 9.59670°, E 123.74990°	JGI-DOE	PacBio	HGAP 2.1.1	4,255,513	13	Scaffold N/L50: 568.3 kbp	53.67	2574179784
22	2141T	*Thiosocius teredinicola* PMS-2141T.STBD.0c.01a	Sulfur-oxidizing	*Kuphus polythalamius*	Kalamansig, Sultan Kudarat, Philippines	N 6.53631°, E 124.048365°	JGI-DOE	PacBio	HGAP 2.0.1	4,790,451	1 (closed circular)	Not applicable	60.08	2751185674

				specimen PMS-2133X										
23	2719K	*Thiosocius* sp. PMS-2719K.STB50.0a.01	Sulfur-oxidizing	*Dicyathifer mannii* specimen PMS-2715W	Infanta, Quezon, Philippines	N 14.68367°, E 121.63690°	JGI-DOE	PacBio	HGAP 2.0.1	5,077,565	1 (closed circular)	Not applicable	58.55	2574179721
24	Ga0198945	*Agarilytica rhodophyticola* strain 017	Agarolytic	Associated with the seaweed *Gracilaria blodgettii*	Lingshui County, Hainan, China	N 18.40828° , E 110.0623°	JGI-DOE	Illumina and PacBio	SOAPdenovo 2.04; Celera Assembler 8.0	6,878,829	1 (closed circular)	Not applicable	40.97	2751185671