Secondary metabolism in the gill microbiota of shipworms (Teredinidae) as revealed by comparison of metagenomes and nearly complete symbiont genomes

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

*authors contributed equally, author order was determined alphabetically
**co-corresponding authors

Abstract

Shipworms play critical roles in recycling wood in the sea. Symbiotic bacteria supply enzymes that the organisms need for nutrition and wood degradation. Some of these bacteria have been grown in pure culture and have the capacity to make many secondary metabolites. However, little is known about whether such secondary metabolite pathways are represented in the symbiont communities within their hosts. In addition, little has been reported about the patterns of host-symbiont co-occurrence. Here, we collected shipworms from the United States, the Philippines, and Brazil, and cultivated symbiotic bacteria from their gills. We analyzed sequences from 22 shipworm gill metagenomes from seven shipworm species and from 23 cultivated symbiont isolates. Using (meta)genome sequencing, we demonstrate that the cultivated isolates represent all the major bacterial symbiont species and strains in shipworm gills. We show that the bacterial symbionts are distributed among shipworm hosts in consistent, predictable patterns. The symbiotic bacteria encode many biosynthetic gene cluster families (GCFs) for bioactive secondary metabolites, only <5% of which match previously described biosynthetic pathways. Because we were able to cultivate the symbionts, and

sequence their genomes, we can definitively enumerate the biosynthetic pathways in these symbiont communities, showing that ~150 out of ~200 total biosynthetic gene clusters (BGCs) present in the animal gill metagenomes are represented in our culture collection. Shipworm symbionts occur in suites that differ predictably across a wide taxonomic and geographic range of host species, and collectively constitute an immense resource for the discovery of new biosynthetic pathways to bioactive secondary metabolites.

Importance

We define a system in which the major symbionts that are important to host biology and to the production of secondary metabolites can be cultivated. We show that symbiotic bacteria that are critical to host nutrition and lifestyle also have an immense capacity to produce a multitude of diverse and likely novel bioactive secondary metabolites that could lead to the discovery of drugs, and that these pathways are found within shipworm gills. We propose that, by shaping associated microbial communities within the host, the compounds support the ability of shipworms to degrade wood in marine environments. Because these symbionts can be cultivated and genetically manipulated, they provide a powerful model for understanding how secondary metabolism impacts microbial symbiosis.

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 (Fig. 1) (3-6). Other shipworms 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 (Fig. 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 byproducts of cellulolytic gut microbes as found in other symbioses. Shipworm symbionts are also essential for nitrogen fixation that helps to offset the low nitrogen content of wood (10, 11). 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, 12, 13). This led to the discovery that these bacteria are exceptional sources of secondary metabolites (14). 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. (13-16). 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.

An early analysis of the *turnerae* T7901 genome revealed nine complex polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) BGCs (14). 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 (17). A second BGC synthesizes the borated polyketide tartrolons D/E, which are antibiotic and potently antiparasitic compounds (18). 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 shipworm ecology.

T. turnerae T7901 is just one of multiple strains and species of γ-proteobacteria living intracellularly in shipworm gills (3, 12), and thus these analyses just begin to describe shipworm secondary metabolism. Many shipworm species are generalists, consuming wood from a variety of sources (1, 19). 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 (20, 21). 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 (19). 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 (22). Other shipworms are found in solid rock and in seagrass (23, 24). 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 (14, 16), the capacity of other shipworm symbionts is still largely unknown. Moreover, several facts indicate that the BGCs found in cultivated isolates might also be produced by symbionts within 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 (17, 18); an investigation of four isolate genomes and one metagenome that observed shared pathways (25); 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 (26). 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. We selected six species of wood-eating shipworms (*B. thoracites*, *N. reynei*, *Bankia setacea*, *Bankia* sp., D*. mannii*, and *Teredo* sp.), comparing these to a seventh sulfide-oxidizing group, *Kuphus* spp. We compared gill metagenomes from 22 specimens comprising seven animal species with the genomes of 23 cultivated bacteria isolated from shipworms. These isolated bacteria included 22 cellulolytic and sulfur-oxidizing isolates cultivated from shipworm tissue samples. By comparing the gill metagenomes to isolate strain genomes, we demonstrate that the cultivated bacterial genomes accurately represent the genomes of symbionts found in the gills, and we show that they share many of the same secondary metabolic BGCs. Moreover, we show that the members of symbiont communities differ among shipworm species, indicating that surveying more host shipworms will lead to discovery of new BGCs and new bacterial symbionts.

Results and Discussion

Sequencing data. Most of the genomes and metagenomes were obtained in this work and are described here for the first time, or in a few cases previously reported genomes/metagenomes were resequenced/reassembled/reanalyzed (see Methods). Two bacterial genomes, *T. turnerae* T7901 and *T. teredinicola* 2141T, and metagenomes of *K. polythalamius* were previously described (7, 14, 22). The resulting statistics and accession numbers are provided in Table S1 A, while specimen and strain origin, many of which have not been previously reported, are given in Table S1 B. For bacterial strains, six of the circular genomes were closed, while remaining assemblies had between 2-141 scaffolds. Metagenome total assembly sizes ranged from 2.6×10^8-1.3×10^9 bp, with N50s of 860-4530 bp. The larger N50s were obtained with the Philippines specimens sequenced at the University of Utah, while others sequenced elsewhere had comparatively shorter N50s.

151 **Mapping cultivated bacteria to gill metagenomes.** A phylogenetic tree created from the 16S
152 rRNA genes of the cultivated bacteria (Fig. 2A and S1) revealed that the strains are all γ-
153 Proteobacteria. Of these, 21 are from Order *Cellvibrionales,* including 11 strains of *T. turnerae*,
154 and 10 strains of diverse cellulolytic bacteria, most of which have not been previously
155 described. An exception is strain BS02, which was recently formally described as the new
156 species, *Teredinibacter waterburyi* (27). The remaining two strains are from Order *Chromatiales*
157 (*Thiosocius* and allies).
158
159 Further, gANI measurements reinforce the 16S rRNA based phylogenetic tree of sequenced
160 strains (Fig. 2, 3, and S1, Table S2). Previously proposed cut-offs for bacterial species
161 differentiation suggest that bacterial strains with gANI values ≥0.95 are conspecific, although
162 several well-known species have lower gANI values (28). The concatenated *T. turnerae* strains
163 are represented by two groups, exemplified by strains T7901 and T7902 (Fig. S2). Within each
164 group, *T. turnerae* strains have gANI values >0.97, whereas between groups the gANI values are
165 ~0.92. This agrees with and reinforces a previously published observation that *T. turnerae* is
166 comprised of two distinct clades and suggests that these clades may in fact constitute distinct
167 but closely related bacterial species (12). Outside of *T. turnerae*, the strains are much less
168 closely related, with AF x gANI values <0.4 (Figs. 3 and S2), indicating that they are all different
169 at the species level.
170
171 Using metagenomic methods, the bacteria living in gills were grouped into bins that represent
172 individual species of bacteria (Fig. 2B). For example, in *Kuphus* spp., >95% of bacterial reads
173 could be mapped to cultivated isolate strain *T. teredinicola* 2141T. Among the three specimens
174 measured, 14 bins mapped to *T. teredinicola* 2141T (Table S2). None of the other specimens in
175 our study had any match to *T. teredinicola* 2141T with gANI >0.90. Normalized by length, these
176 bins had a total gANI = 0.96 (Table 1). In comparison to values obtained in the phylogenetic
177 tree, these data suggest that *T. teredinicola* 2141T is conspecific with the uncultivated
178 symbionts in the metagenomes of *Kuphus* spp.
179
180 Similarly, *Cellvibrionaceae* strain 2753L was mapped to 20 bins in *D. mannii* and *B. thoracites*
181 specimens, with a total gANI >0.99. When bins were mapped to discrete strains as shown in Fig.
182 2B, the gANI was 0.96-0.99 to a single strain, with much lower identity to other strains
183 sequenced. These data demonstrate a high level of identity between cultivated isolates and the
184 strains present within shipworm gills, suggesting that in some cases these are near identical
185 strains to those present within the shipworms.
186
187 In other cases, either because we had multiple strains representing a species (as in *T. turnerae*)
188 or because the identity to single strains was not as pronounced, we described bins as "*T.*
189 *turnerae*", "Genus Teredinibacter", and "Family Cellvibrionaceae". These still had relatively high
190 identities to cultivated isolates. For example, the *Teredo* sp. bins in total had a gANI > 0.98 to
191 cultivated isolates in our strain collection. It is likely that the metagenomes from these animals
192 were not as similar to cultivated isolates because, in those cases, we compared isolates from
193 Philippines specimens with metagenomes of Brazilian animals.
194

195 In sum, these data demonstrate conclusively that the cultivated isolates obtained from
196 shipworm gills accurately represent the strains found within shipworms. The data suggest that
197 the isolates are the same species as the true symbionts found within the animals, and in many
198 cases they are >99% DNA sequence identical at the whole genome level. The data reveal that
199 >85% of the DNA in each specimen's gill metagenome is represented by a cultivated isolate in
200 our collection (with the exception of one specimen), and that the remaining <15% of the DNA
201 belongs to multiple, low-abundance species, most of which are not reproducibly found in
202 multiple shipworm isolates. Further, much of the shipworm literature focuses on the readily
203 cultivable *T. turnerae*. We show that *T. turnerae* is dominant in some species, but it is very
204 minor or even absent in others.
205
206 **Strain variation increases genetic diversity of shipworm microbiota.** Metagenome binning
207 defined the major symbiont species present in shipworm gills to be a relatively simple mixture
208 of one to three species. Since we had deep sequencing of the major metagenomic bacterial
209 species, we expected to be able to provide complete assemblies. In other instances, using
210 similarly deep data, we have been able to obtain relatively complete assemblies, or even to
211 assemble whole bacterial genomes from metagenomes (29). However, our metagenome bin
212 N50s were only in the very low thousands.
213
214 When investigating the causes underlying the challenge of assembly, we noted that we often
215 obtained very similar contigs with different copy numbers. For example, a single metagenome
216 bin containing BCS2-like contigs is shown in Table S3. Pairwise identities between contigs are
217 very high, between 93-98% DNA sequence identical, indicating that these bins are comprised of
218 mixtures of very closely related bacteria. We saw a very similar phenomenon in a recent
219 investigation of *K. polythalamius* symbionts (7). In that case, the strains were nearly identical
220 and could not be resolved by 16S rRNA gene sequences, which were 100% identical. Thus, we
221 developed a different method to quantify strain-level variation that was observed using
222 metagenomics.
223
224 In the *Kuphus* study, we cut the DNA gyrase B gene into 50 bp segments and aligned single
225 reads to each 50 bp segment (7). By quantifying reads for each observed SNP, we confirmed
226 that the gill symbiont species consisted of several strains, and we quantified their relative
227 abundances. Here, we expanded this previous knowledge by investigating the major strains
228 found in the remaining shipworm species, using the same method. We show four additional
229 examples (Fig. S3) in which we can quantify multiple strains of each different bacterial symbiont
230 species, but the same phenomenon obtains in all of the metagenomes. This analysis shows that
231 similar strain variation is a widespread phenomenon in shipworm gills, and not just restricted to
232 *K. polythalamius*. We believe that strain variation is likely to be an important source of BGC
233 variation, as described further below.
234
235 **Discovery and analysis of BGCs.** Knowing that the cultivated bacteria represent the major
236 symbiont species present in gill metagenomes, we next compared secondary metabolism
237 between these specimens and isolates. To start, we took an inventory of the BGC content in our
238 assembled sequences. Analysis using antiSMASH (30) revealed a large number of BGCs: 431

239 BGCs were identified in the 23 cultivated isolates alone. Because raw antiSMASH output
240 includes many hypothetical or poorly characterized BGCs, we chose to focus on well-
241 characterized classes of secondary metabolic proteins and pathways: polyketide synthases
242 (PKSs), nonribosomal peptide synthetases (NRPSs), siderophores, terpenes, homoserine
243 lactones, and thiopeptides. Using these criteria, we identified 168 BGCs from 23 cultivated
244 isolates and 401 BGCs from 22 shipworm gill metagenomes (Fig. 4). Because the genomes of
245 cultivated isolates were well assembled, we could discern and analyze entire BGCs. By contrast,
246 animal metagenomes had smaller contigs, so that BGCs were fragmented.
247
248 The BGCs identified in this study nearly universally originate from Order *Cellvibrionales*, with
249 very few BGCs found in the sulfide oxidizing strains *Chromatiales*. Thus, the cellulolytic
250 shipworm symbionts are rich sources of diverse BGCs. We found only five BGCs that were
251 similar to previously identified clusters from outside of shipworms, based upon >70% of genes
252 conserved in antiSMASH. The remainder appeared to be unknown or uncharacterized BGCs. In
253 turn, the new BGCs are likely to represent new compounds, while characterized BGCs represent
254 those for previously identified compounds. In addition, it is possible that some of the new BGCs
255 may represent known compounds, for which biosynthetic pathways have not yet been
256 discovered. This result further supports a previous analysis comparing genomes across domain
257 Bacteria, which revealed that *T. turnerae* represents a notably rich, yet nearly untapped, source
258 of new secondary metabolite genes (16).
259
260 To facilitate comparison between metagenomes, we grouped all 569 BGCs into 122 gene
261 cluster families (GCFs), where each GCF is comprised of closely related BGCs (31, 32) (Fig. 5 and
262 Table S4). BGCs grouped into a single GCF are highly likely to encode the production of identical
263 or closely related secondary metabolites.
264
265 Some important BGCs were excluded using our method. For example, we analyzed the genome
266 of *Chromatiales* strain 2719K and discovered a gene cluster for tabtoxin (33, 34) or a related
267 compound (Fig. 6). This cluster does not contain common PKS/NRPS elements and thus is not
268 one of the GCFs shown in Figures 5, 7, or 8. A key biosynthetic gene in the tabtoxin-like cluster
269 was pseudogenous in strain 2719K, but the *D. mannii* gill metagenome contained an apparently
270 functional pathway. Tabtoxin is an important β-lactam that is used by *Pseudomonas* in plant
271 pathogenesis (35, 36).
272
273 **Comparison of isolate and gill BGCs.** Of 401 BGCs identified in the metagenomes, 305 of them
274 also had close relatives in cultivated isolates, indicating that ~75% of BGCs in the metagenomes
275 are covered in our sequenced culture collection (Fig. 4). Conversely, of 168 isolate BGCs, 148
276 (90%) of them are found in the metagenomes. Thus, sequencing additional cultivated isolates in
277 our strain collections is likely to yield additional novel BGCs. Since the 11 *T. turnerae* strains
278 analyzed in this project contain different BGCs, we speculate that the additional BGC variation
279 is due to the observed strain variation in the shipworm gills.
280
281 It is notoriously difficult to quantify BGCs in metagenomes, which usually contain relatively
282 small contigs. Since BGCs in the classes that we analyzed are usually between 10 kbp to >100

283 kbp in length, each BGC is usually represented by multiple, short contigs, which are not easily
284 mapped. Here, we had an advantage in that the cultivated isolates accurately represented the
285 gill metagenomes: we could map the identified metagenomic contigs to the assembled BGCs
286 found in cultivated isolates.
287
288 Using this mapping, we could accurately estimate the number of unique BGCs in the gill
289 symbiont community. For example, 305 metagenome BGCs are synonymous with 148 isolate
290 BGCs, indicating that the metagenome BGC count can be estimated to be approximately double
291 the actual number of BGCs. To verify this estimate, we selected GCFs 2, 3, 5 and 8, aligning
292 metagenomic contigs against the BGCs from cultivated isolates (Fig. S4). In the metagenomes,
293 out of the 401 total BGCs identified, 100 were members of these four GCFs, but some of them
294 were just fragments of the full-length BGCs found in cultivated isolates. When the 100
295 metagenomic BGCs were aligned to their congeners in cultivated isolates, they could be
296 collapsed into 46 unique BGCs. Thus, using two different approaches, we could accurately
297 estimate that the 401 metagenomic BGCs of all GCFs represent ~200 actual BGCs in the
298 shipworm gills. To the best of our knowledge, this type of estimate has not been possible for
299 other metagenomes/symbioses and represents a uniquely powerful aspect of this system.
300
301 Only 8 GCFs are widely distributed in 10 or more isolates, and these are mostly pathways that
302 are universal or nearly universal in *T. turnerae*, which is overrepresented in our data set (Figs. 7
303 and 8). By contrast to isolate genomes in which we found many GCFs that occur in only a single
304 genome, in the metagenomes most of the 107 GCFs are found in multiple specimens. Forty-five
305 GCFs are found in multiple species of shipworms. Sixty-two GCFs were only found in a single
306 shipworm species; 26 of these were only found in a single specimen (Fig. 5). These data
307 demonstrate that accessing diverse shipworm specimens, as well as diverse shipworm species,
308 will lead to the discovery of many novel BGCs. In addition, this result reinforces the strain-level
309 variation found in shipworms revealed both in metagenome assembly results as well as in DNA
310 gyrase B SNP analysis.
311
312 To obtain a more refined view of BGC distribution, we first used the MultiGeneBlast (31) output
313 to construct a similarity network (Fig. 7). The network provided an easily interpretable diagram
314 of how GCFs are distributed among bacteria. However, a notable shortcoming was observed. In
315 a long-term drug discovery campaign, we have found the tartrolon BGC in nearly all *T. turnerae*
316 strains ((18), unpublished observation). However, this BGC was observed in only a few of the *T.*
317 *turnerae*-hosting shipworms via MultiGeneBlast. This is caused by a technical problem in
318 assembly that we often see with large *trans*-acyltransferase (*trans*-AT) pathways from complex
319 samples (37). Thus, we were concerned that networking might underreport the similarity of
320 some types of biosynthetic pathways.
321
322 To remedy this problem, we obtained GCFs from cultivated isolates and searched them against
323 metagenome contigs using tBLASTn (Fig. 8). This provided an orthogonal view of secondary
324 metabolism in shipworms, revealing the presence of the tartrolon pathway, as well as other
325 pathways that do not assemble well in metagenomes because of characteristics such as
326 repetitive DNA sequences. A weakness of this second method is that it does not tell us whether

327 two pathways are related enough to encode the production of similar compounds. Thus, these
328 two methods provide different insights into BGCs in shipworm gills.
329
330 The close similarity of BGCs between cultivated isolates and metagenomes further reinforced
331 the species identities determined by gANI (Table 1). Since secondary metabolism is often one of
332 the most variable genomic features in bacteria, the sharing of multiple pathways between gills
333 and isolates provides further evidence that the isolates are representative of the true
334 symbionts found in gills.
335
336 We identified three categories of GCFs: (a) GCFs that are widely shared among shipworm
337 species, (b) GCFs that were specific to select shipworm and symbiont species, and (c) GCFs that
338 were distributed among specimens without obvious relationship to host or symbiont species
339 identity. These pathways are described in the following sections.
340
341 **(a) Widely shared GCFs.** Four pathways (GCF_2, GCF_3, GFC_5, and GCF_8) were prevalent in
342 all wood-eating shipworms, regardless of sample location (Figs. 7 and 8). These GCFs were
343 encoded in the genomes of *T. turnerae,* the most widely distributed shipworm symbiont, and
344 those of several other *Cellvibrionales* symbiont isolates from wood eating shipworms (especially
345 the pathway-rich isolate 2753L).
346
347 The most widely occurring pathway in shipworm gill metagenomes is GCF_3. It was identified in
348 all gill metagenomes with cellulolytic symbionts, including the metagenomes of specimen *B.*
349 *setacea* BSG2. It occurs in all *T. turnerae* strains, as well as in *Cellvibrionales* strains 2753L and
350 Bs08. It was first annotated as "region 1" in the *T. turnerae* T7901 genome and encodes an
351 elaborate hybrid *trans*-AT PKS-NRPS pathway (14). Unlike all other GCFs identified in shipworm
352 metagenomes and isolates, GCF_3 could be subdivided into at least three discrete categories,
353 each of which included different gene content (Fig. 9). The first category, identified in *T.*
354 *turnerae* T7901, encodes a PKS and a single NRPS, in addition to several potential modifying
355 enzymes. In strain Bs08, instead of just a single NRPS, GCF_3 contains three NRPS genes.
356 Presumably, Bs08 and T7901 produce products with similar or identical polyketides and amino
357 acids, except that BS08 adds two more amino acids to the chain. *Cellvibrionales* 2753L encoded
358 the third pathway type, which was similar to that found in T7901 except with different flanking
359 genes that might encode modifying enzymes. Thus, T7901 and 2753L might make identical or
360 very similar polyketide-peptide scaffolds, which are modified slightly differently after scaffold
361 synthesis. The presence of a single GCF that encodes similar but non-identical products
362 suggests a dynamic pathway evolution within shipworms.
363
364 GCF_2 encodes a NRPS / *trans*-AT PKS pathway, the chemical products of which are unknown. It
365 is found in all shipworm specimens in this study and in all *T. turnerae* strains. It is also present in
366 *Cellvibrionales* strain 2753L. This explains its presence in *B. thoracites* despite the absence of *T.*
367 *turnerae* in this species. GCF_2 is synonymous with "region 3" described in the annotation of
368 the *T. turnerae* T7901 genome (14).
369

370 GCF_5 encodes a combination of terpene cyclase and predicted arylpolyene biosynthetic genes,
371 which were unrecognized in the initial sequence analysis of *T. turnerae* T7901 (14), since the
372 arylpolyene pathways are recent discoveries (38). Although the cyclase and surrounding regions
373 have all of the genes necessary to make and export hopanoids, the GCF_5 biosynthetic product
374 is unknown. In addition to occurring in all *T. turnerae* strains, GCF_5 is present in *Cellvibrionales*
375 strains 1120W and 2753L. The pathway was detected in all wood-eating specimens except
376 *Teredo* sp. TBF07 (Fig. 9).
377
378 GCF_8 is exemplified by the previously described turnerbactin BGC, from *T. turnerae* T7901.
379 Turnerbactin is a catecholate siderophore, crucial to iron acquisition in *T. turnerae* (17). The
380 BGC for turnerbactin was identified and described as "region 7" in the previously published *T.*
381 *turnerae* T7901 genome. GCF_8 is present in all *T. turnerae* genomes sequenced here. Other
382 *Cellvibrionales* strains, including 2753L from *B. thoracites* and Bs08 from *B. setacea (*neither of
383 which contains *T. turnerae)*, also encode turnerbactin-like siderophore synthesis. GCF_8 was
384 also found in the metagenome of one specimen of *B. thoracites*. Beyond bacterial iron
385 acquisition, siderophores are also important in strain competition and potentially in host animal
386 physiology (39, 40), possibly explaining the widespread distribution of GCF_8. From the
387 clustering pattern in Fig. 7, it is likely that GCF_8 comprises at least three different, but related
388 types of gene clusters. Thus, GCF_8 likely represents catecholate siderophores, but not
389 necessarily turnerbactin.
390
391 **(b) Bacterial species-specific GCFs.** In addition to the four GCFs described above that have a
392 wide distribution, GCFs 1, 4, and 11 were found in all *T. turnerae*-containing shipworms. GCF_1
393 is a *trans*-AT PKS-NRPS pathway that appears to be split into two clusters in some shipworm
394 isolates, including *T. turnerae* T7901, in which it was previously annotated as "region 4" and
395 "region 5". GCF_4 is the previously described "region 8" PKS-NRPS from *T. turnerae* T7901.
396 Most notably, GCF_11 encodes tartrolon biosynthesis (18). Tartrolon is an antibiotic and potent
397 antiparasitic agent isolated from culture broths of *T. turnerae* T7901 (18, 41, 42). It has also
398 been identified in the cecum of the shipworm. It was proposed that the bacteria synthesize
399 tartrolon in the gill, and it is transferred to the cecum where it may play a role in keeping the
400 digestive tract free of bacteria (18).
401
402 The gill metagenomes of *D. mannii* and *B. thoracites* indicate the abundant presence of 2753L-
403 like strains. Like *T. turnerae* T7901, the 2753L isolate genome encodes GCFs 2, 3, and 5.
404 However, 2753L contains several GCFs not found in *T. turnerae*, including GCFs 6, 10, 12, 13, 14,
405 16, 30, and 31 (listed in order of their relative frequency of occurrence in samples). All of these
406 GCFs are also evident in *D. mannii* and *B. thoracites* gill metagenomes. These are PKS and NRPS
407 clusters that lack close relatives according to antiSMASH annotation and thus have a potential
408 to synthesize novel secondary metabolite classes.
409
410 Brazilian shipworms *Bankia* sp. and *Teredo* sp. contain *T. turnerae* and the major pathways
411 found in *T. turnerae*, but they are dominated by symbiont genomes from other symbiotic
412 *Cellvibrionaceae* bacteria. Although those species are not represented in our current culture
413 collection, they are closely related to isolate 1162T from a Philippine specimen

414 of *Lyrodus* sp. The metagenomes of *Bankia* sp. and *Teredo* sp. contain many GCFs that are not
415 found in sequenced isolates (Fig. S5B). In addition, the GCFs found in *Bankia* sp. and *Teredo* sp.
416 are not completely overlapping, implying that the *Cellvibrionaceae* bacteria found in these
417 different host species are distinct. A high number of GCFs were found, indicating that
418 potentially the symbionts might have a similar GCF content as the GCF-rich isolate 2753L.
419
420 The *B. setacea* specimens shown in Fig. S5B contain pathways specifically found in
421 *Cellvibrionaceae* isolate BSC2, which is the major bacterium observed in the *B. setacea* gill
422 metagenome sequences.
423
424 The *K. polythalamius* gill metagenome and its cultivated sulfur oxidizing symbiont *T.*
425 *teredinicola* contain relatively few BGCs, but strikingly two NRPS-containing GCFs have been
426 found in all shipworm specimens containing the sulfide-oxidizing symbionts (*K. polythalamius*
427 and *D. mannii*) and all sulfide-oxidizing symbiont isolates (*T. teredinicola* and isolate 2719K).
428 One of these, GCF_17, is shown in Fig. 8. Based on our analyses, it is clear that the cellulolytic
429 symbionts contain more abundant and diverse BGCs.
430
431 **(c) GCFs for which patterns of occurrence are not obviously related to host species identity.**
432 Overall, the most abundant pathways in shipworms were identical to those from the cultivated
433 isolate genomes that were mapped to each shipworm metagenome (Figs. 7 and 8). Since
434 specific bacterial symbionts are distributed among shipworm hosts in patterns that are
435 predicted by host species identity and life habits, the presence of abundant GCFs also follow
436 similar patterns. However, as described above, many pathways were found only once or
437 occurred relatively rarely among symbiont genomes and gill metagenomes. In these cases,
438 trends of host symbiont co-occurrence could not be discerned. This trend is reinforced in Fig. 7,
439 where most GCFs in the diagram occur only once (single, unlinked spots). Thus, while the
440 occurrence of several biosynthetic pathways is evolutionarily conserved among host species
441 and thus likely have a uniquely critical role in the symbiosis, most are not conserved. These
442 observations suggest that more comprehensive sampling of shipworm specimens, species, and
443 cultivated isolates will yield many additional, unanticipated BGCs.
444
445 **Variability in conserved shipworm GCFs increases potential compound diversity.** Even among
446 conserved GCFs, some variability was observed. This is evident in the BGC network analysis
447 shown in (Fig. 7), where subclusters indicate slightly different GCF organization. For example, in
448 the ubiquitous GCF_3, the three different pathway variants appear in the network as bulges
449 within the cluster. The siderophore pathway GCF_8 contains one central cluster, encoding
450 turnerbactin pathways, and an extended arm that appears to encode compounds that are
451 related to, but not identical to, turnerbactin. Thus, the shape of the network clusters indicates
452 the potential chemical diversity encoded in individual GCFs.
453
454 **Conclusions**
455
456 In shipworms, cellulolytic bacteria were long known to specifically inhabit gills and were
457 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. (14, 16) – led us to investigate shipworms as a source of new bioactive compounds.

Here, we show that cultivated isolates obtained from shipworm gills accurately represent the bacteria living within the gills. They are the same species, and often are nearly identical at the strain level. They contain many of the same BGCs. The gills of shipworms contain about 1-3 major species of symbiotic bacteria, along with a small percentage of other less consistently occurring bacteria. Complicating this relatively simple picture, there is significant strain variation within shipworms. The observed symbiont species mixtures are representative of the animal lifestyles. For example, *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.

The key finding is that BGCs in the metagenomes are represented in the strains in our culture collection. This is a rare event in the biosynthetic literature. In most other marine systems, it has been very challenging to cultivate the symbiotic bacteria responsible for secondary metabolite production (44). In some organisms, such as humans, there are many representative cultivated isolates that produce secondary metabolites, but connecting those metabolites to human biology, or even to their existence in humans, is quite challenging (16, 45). Here, we have defined an experimentally tractable system to investigate chemical ecology that circumvents these limitations. Our 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.

It has not escaped our notice that this work provides the foundation for understanding the connection between symbiont community composition, secondary metabolite complement, and host lifestyle and ecology. It has proven difficult to link these factors together in relevant models. The existence of aquaculture and transformation methods for shipworms and their symbiotic bacteria will enable a rigorous, hypothesis-driven understanding of the role of complex metabolism in symbiosis.

Methods

Collection and processing of biological material. Shipworm samples (Table S1) 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

502 and identified. Specimens were dissected using a dissecting stereoscope. Taxonomic vouchers
503 (valves, pallets, and siphonal tissue for sequencing host phylogenetic markers), were retained
504 and stored in 70% ethanol. The gill was dissected, rinsed with sterile seawater, and divided for
505 bacterial isolation and metagenomic sequencing. Once the gill was dissected it was processed
506 immediately or flash-frozen in liquid nitrogen.
507
508 Of the animals that we obtained in field collections we analyzed three specimens each of
509 *Bactronophorus thoracites*, *Kuphus* spp., *Neoteredo reynei*, and *Teredo* sp., two specimens of
510 *Bankia* sp., and five specimens of *Bankia setacea*. These animals were divided into three
511 geographical regions (Fig. 1): the Philippines (*B. thoracites* and *D. mannii* from Infanta, Quezon;
512 *Kuphus* spp. from Mindanao and Mabini); Brazil (*N. reynei* from Rio de Janeiro, *Teredo* sp., and
513 *Bankia* sp. from Ceará); and the United States (*B. setacea*). The purpose of sampling this range
514 was to determine whether there are any geographical differences in gill symbiont occurrence.
515 Most of the shipworms were obtained from mangrove wood, with the exception of *B. setacea*
516 from unidentified found wood, and *Kuphus* spp. from both found wood and mud.
517
518 **Bacterial isolation, DNA extraction and analysis.** *Teredinibacter turnerae* strains (with T prefix)
519 were isolated using the method described in Distel el al. 2002 (13), while *Bankia setacea*
520 symbionts (with Bs prefix) were obtained using the technique indicated in O'Connor et al. 2014
521 (9). Sulfur-oxidizing symbionts were isolated using the protocol specified in Altamia et al. 2019
522 (22). For this study, additional *T. turnerae* and novel cellulolytic symbionts from Philippine
523 specimens (with prefix PMS) were isolated (Table S1). Briefly, dissected gill were homogenized
524 in sterile 75% natural seawater buffered with 20 mM HEPES, pH 8.0 using a Dounce
525 homogenizer. Tissue homogenates were either streaked on shipworm basal medium cellulose
526 (5) plates (1.0% Bacto Agar) or stabbed into soft agar (0.2% Bacto Agar) tubes and incubated at
527 25 °C until cellulolytic clearings developed. Cellulolytic bacterial colonies were subjected to
528 several rounds of restreaking to ensure clonal selection. Contents of soft agar tubes with
529 clearings were streaked on fresh cellulose plates to obtain single colonies. Pure colonies were
530 then grown in 6 mL SBM cellulose liquid medium in 16 × 150 mm test tubes until the desired
531 turbidity was observed. For long-term preservation of the isolates, a turbid medium was added
532 to 40% glycerol at 1:1 ratio and frozen at -80 °C. Bacterial cells in the remaining liquid medium
533 were pelleted by centrifugation at 8,000 *g* and then subjected to genomic DNA isolation. The
534 small-subunit ribosomal (SSU) 16S rRNA gene of the isolates was then PCR amplified using 27F
535 (5'-AGAGTTTGATCCTGGCTCAG-3') and 1492R (5'-GGTTACCTTGTTACGACTT-3') from the
536 prepared genomic DNA and sequenced. Phylogenetic analyses of 16S rRNA sequences was
537 performed using programs implemented in Geneious, version 10.2.3. Briefly, sequences were
538 aligned using MAFFT (version 7.388) by using the E-INS-i algorithm. The aligned sequences were
539 trimmed manually, resulting in a final aligned dataset of 1,125 nucleotide positions.
540 Phylogenetic analysis was performed using FastTree (version 2.1.11) using the GTR substitution
541 model with optimized Gamma20 likelihood and rate categories per site set to 20.
542
543 Genomic DNA used for whole genome sequencing of novel isolates and select *T. turnerae*
544 strains were prepared using CTAB/phenol/chloroform DNA extraction method detailed in
545 https://www.pacb.com/wp-content/uploads/2015/09/DNA-extraction-chlamy-CTAB-JGI.pdf.

546 The purity of the extracted genomic DNA was then assessed spectrophotometrically using
547 Nanodrop and the quantity was estimated using agarose gel electrophoresis. Samples that
548 passed the quality control steps were submitted to Joint Genome Institute – Department of
549 Energy (JGI-DOE) for whole genome sequencing. The sequencing platform and assembly
550 method used to generate the final isolate genome sequences used in this study are detailed in
551 Table S1 A.
552
553 **Metagenomic DNA extraction.** Gill tissue samples from Philippine shipworm specimens (Table
554 S1 B) were flash-frozen in liquid nitrogen and stored at -80°C prior to processing. Bulk gill
555 genomic DNA was purified by Qiagen Blood and Tissue Genomic DNA Kit using the
556 manufacturer's suggested protocol.
557
558 Gill tissue samples from Brazil shipworm specimens were pulverized by flash-freezing in liquid
559 nitrogen and submitted to metagenomic DNA purification by adapting a protocol previously
560 optimized for total DNA extraction from cnidaria tissues (46, 47). Briefly, shipworms gills were
561 carefully dissected (taking care not to get intersections with other organs), submitted to a
562 series of five washes with 3:1 sterile seawater / distilled water for removal of external
563 contaminants, and macerated until powdered in liquid nitrogen. Powdered tissues (~150 mg)
564 were then transferred to 2 mL microtubes containing 1 mL of lysis buffer [2% (m/v)
565 cetyltrimethyl ammonium bromide (Sigma Aldrich), 1.4 M NaCl, 20 mM EDTA, 100 mM Tris-HCl
566 (pH 8.0), with freshly added 5 µg proteinase K (v/v; Invitrogen), and 1% 2-mercaptoethanol
567 (Sigma Aldrich)] and submitted to five freeze-thawing cycles (-80 °C to 65 °C). Proteins were
568 extracted by washing twice with phenol:chloroform:isoamyl alcohol (25:24:1) and once with
569 chloroform. Metagenomic DNA was precipitated with isopropanol and 5 M ammonium acetate,
570 washed with 70% ethanol, and eluted in TE buffer (10 mM Tris-HCl, 1 mM EDTA). Metagenomic
571 libraries were prepared using the Nextera XT DNA Sample Preparation Kit (Illumina) and
572 sequenced with 600-cycle (300 bp paired-end runs) MiSeq Reagent Kits v3 chemistry (Illumina)
573 at the MiSeq Desktop Sequencer.
574
575 **Metagenome sequencing and assembly.** Five *Bankia setacea* metagenome sequencing raw
576 read files were obtained from the JGI database and reassembled using the methods described
577 below (for accession numbers, see Table S1 A). *Kuphus polythalamius* gill metagenomes
578 (KP2132G and KP2133G) were obtained from a previous study (7). Metagenomes from *Kuphus*
579 sp. specimen KP3700G and *Dicyathifer mannii* and *Bactronophorus thoracites* specimens were
580 sequenced using an Illumina HiSeq 2000 sequencer with ~350 bp inserts and 125 bp paired-end
581 runs at the Huntsman Cancer Institute's High Throughput Genomics Center at the University of
582 Utah. Illumina fastq reads were trimmed using Sickle (48) with the parameters (pe sanger -q 30
583 –l 125). The trimmed FASTQ files were converted to FASTA files and merged using the Perl
584 script 'fq2fq' in IBDA_ud package (49). Merged FASTA files were assembled using IDBA_ud with
585 standard parameters in the Center for High Performance Computing at the University of Utah.
586 For metagenome samples from Brazil, all *Neoterdo reynei* gill metagenomic samples previously
587 analyzed were re-sequenced here to improve coverage depth (26). *Teredo* sp. and *Bankia* sp.
588 gill metagenomes were sequenced using Illumina Miseq. The raw reads were assembled using
589 either the metaspades pipeline of SPAdes (50, 51) or IDBA-UD (49). Before assembly, raw reads

590 were merged using BBMerge (52). Non-merged reads were filtered and trimmed using FaQCs
591 (53).
592
593 **Identification of bacterial sequences in metagenomic data.** Assembly-assisted binning was
594 used to sort and analyze trimmed reads and assembled contigs into clusters putatively
595 representing single genomes using MetaAnnotator (54). Each binned genome was retrieved
596 using Samtool (55, 56). To identify bacterial genomes, genes for each bin were identified with
597 Prodigal (57). Protein sequences for bins with coding density >50% were searched against NCBI
598 nr database with DIAMOND (58). Bins with 60% of the genes hitting bacterial subject in the nr
599 database were considered to originate from bacteria.
600
601 For *B*. *setacea* metagenome samples and the ones from Brazil, structural and functional
602 annotations were carried out using DFAST (59), including only contigs with length ≥500 bp. All
603 metagenomes were binned using Autometa (60). First, each contig's taxonomic identity was
604 predicted using make_taxonomy_table.py, including only contigs ≥1000 bp. Predicted bacterial
605 and archaeal contigs were binned (with recruitment via supervised machine learning) using
606 run_autometa.py.
607
608 **gANI comparison and reads counts calculation.** Each bacterial bin was compared to the 23
609 shipworm isolate genomes using gANI and AF values (61). With a cut-off of AF >0.5 and gANI
610 >0.9, the bacterial bins from each metagenome were mapped to cultivated bacterial genomes,
611 and cultivated bacterial genomes were mapped against each other (Table S2). The major but
612 not mapped bins in each genome were classified using gtdb-tk (62). The read counts for each
613 mapped bin were either retrieved from MetaAnnotator output or calculated using bbwrap.sh
614 (sourceforge.net/projects/bbmap/) with the parameters: kfilter=22 subfilter=15 maxindel=80.
615
616 **Building BGC similarity networks.** BGCs were predicted from the bacterial contigs of each
617 metagenome and from cultivated bacterial genomes using antiSMASH 4.0 (30). From the
618 predictions, only BGCs for PKSs, NRPSs, siderophores, terpenes, homoserine lactones, and
619 thiopeptides (as well as combinations of these biosynthetic enzyme families) were included in
620 succeeding analyses. An all-versus-all comparison of these BGCs was performed using
621 MultiGeneBlast (31) following the protocol previously reported (63). Bidirectional
622 MultiGeneBlast BGC-to-BGC hits were considered to be reliable. In metagenome data, some
623 truncated BGCs only showed single-directional correlation to a full length BGC. Those single-
624 directional hits were refined as follows: protein translations of all coding sequences from the
625 BGCs were compared in an all-versus-all fashion using blastp search. Only protein hits that had
626 at least 60% identity and at least 80% coverage to both query and subject were considered as
627 valid hits. A single-directional MultiGeneBlast BGC-to-BGC hit was retained if there were at
628 least *n*-2 number of proteins (*n* is the number of proteins in the truncated BGC) passing the
629 blastp refining. The remaining MultiGeneBlast hits were used to construct a network in
630 Cytoscape (64). Finally, each BGC cluster (GCF) that had relative low number of bidirectional
631 correlations were manually checked by examining the MultiGeneBlast alignment.
632

633 **Occurrence of GCFs in metagenomes.** Based on the GCFs identified in previous step, the core
634 biosynthetic proteins from each GCF were extracted and queried (NCBI tblastn) against each
635 metagenome assembly. A threshold of query coverage of >50% and identity > 90% was applied
636 to remove the nonspecific hits, and the remaining hits, in combination with the MultiGeneBlast
637 hits, were used to make the matrix of GCFs occurrence in metagenomes.
638
639 **Acknowledgments.** All collections followed Nagoya Protocol requirements; Brazilian sampling
640 were performed under SISBIO license number 48388, and genetic resources accessed under the
641 authorization of the Brazilian National System for the Management of Genetic Heritage and
642 Associated Traditional Knowledge (SisGen permit number A2F0DA0). We thank the Genomics
643 and Bioinformatics Center of Drug Research and Development Center of Federal University of
644 Ceara for technical support.
645 The work was completed under supervision of the Department of Agriculture-Bureau of
646 Fisheries and Aquatic Resources, Philippines (DA-BFAR) in compliance with all required legal
647 instruments and regulatory issuances covering the conduct of the research. All Philippine
648 specimens were collected under Gratuitous Permit numbers FBP-0036-10, GP-0054-11, GP-
649 0064-12, GP-0107-15, and GP-0140-17. We thank the governments and municipalities of the
650 Philippines and Brazil for access and help.
651 This work was also supported by the National Council of Technological and Scientific
652 Development (CNPq) (http://cnpq.br) and by the Coordination for the Improvement of Higher
653 Education Personnel (CAPES) (http://www.capes.gov.br) under the grant numbers
654 473030/2013-6 and 400764/2014-8 to AETS
655 Research reported in this publication was supported by the Fogarty International Center of the
656 National Institutes of Health under Award Number U19TW008163. The content is solely the
657 responsibility of the authors and does not necessarily represent the official views of the
658 National Institutes of Health. The work was supported in part by US NOAA OER award
659 #NA190AR0110303
660
661
662 **References**
663
664 1. Distel DL, Amin M, Burgoyne A, Linton E, Mamangkey G, Morrill W, Nove J, Wood N,
665 Yang J. 2011. Molecular phylogeny of Pholadoidea Lamarck, 1809 supports a single
666 origin for xylotrophy (wood feeding) and xylotrophic bacterial endosymbiosis in Bivalvia.
667 Mol Phylogenet Evol 61:245-54.
668 2. Turner RD. 1966. A survey and illustrated catalogue of the Teredinidae (Mollusca:
669 Bivalvia). Harvard University Press, Cambridge.
670 3. Distel DL, Beaudoin DJ, Morrill W. 2002. Coexistence of multiple proteobacterial
671 endosymbionts in the gills of the wood-boring Bivalve Lyrodus pedicellatus (Bivalvia:
672 Teredinidae). Appl Environ Microbiol 68:6292-9.
673 4. Luyten YA, Thompson JR, Morrill W, Polz MF, Distel DL. 2006. Extensive variation in
674 intracellular symbiont community composition among members of a single population
675 of the wood-boring bivalve Lyrodus pedicellatus (Bivalvia: Teredinidae). Appl Environ
676 Microbiol 72:412-7.

677 5. Waterbury JB, Calloway CB, Turner RD. 1983. A cellulolytic nitrogen-fixing bacterium
678 cultured from the gland of deshayes in shipworms (bivalvia: teredinidae). Science
679 221:1401-3.
680 6. Ekborg NA, Morrill W, Burgoyne AM, Li L, Distel DL. 2007. CelAB, a multifunctional
681 cellulase encoded by Teredinibacter turnerae T7902T, a culturable symbiont isolated
682 from the wood-boring marine bivalve Lyrodus pedicellatus. Appl Environ Microbiol
683 73:7785-8.
684 7. Distel DL, Altamia MA, Lin Z, Shipway JR, Han A, Forteza I, Antemano R, Limbaco M,
685 Tebo AG, Dechavez R, Albano J, Rosenberg G, Concepcion GP, Schmidt EW, Haygood
686 MG. 2017. Discovery of chemoautotrophic symbiosis in the giant shipworm Kuphus
687 polythalamia (Bivalvia: Teredinidae) extends wooden-steps theory. Proc Natl Acad Sci U
688 S A 114:E3652-E3658.
689 8. Betcher MA, Fung JM, Han AW, O'Connor R, Seronay R, Concepcion GP, Distel DL,
690 Haygood MG. 2012. Microbial distribution and abundance in the digestive system of five
691 shipworm species (Bivalvia: Teredinidae). PLoS One 7:e45309.
692 9. O'Connor RM, Fung JM, Sharp KH, Benner JS, McClung C, Cushing S, Lamkin ER,
693 Fomenkov AI, Henrissat B, Londer YY, Scholz MB, Posfai J, Malfatti S, Tringe SG, Woyke
694 T, Malmstrom RR, Coleman-Derr D, Altamia MA, Dedrick S, Kaluziak ST, Haygood MG,
695 Distel DL. 2014. Gill bacteria enable a novel digestive strategy in a wood-feeding
696 mollusk. Proc Natl Acad Sci U S A 111:E5096-104.
697 10. Lechene CP, Luyten Y, McMahon G, Distel DL. 2007. Quantitative imaging of nitrogen
698 fixation by individual bacteria within animal cells. Science 317:1563-6.
699 11. Charles F, Sauriau PG, Aubert F, Lebreton B, Lantoine F, Riera P. 2018. Sources
700 partitioning in the diet of the shipworm Bankia carinata (J.E. Gray, 1827): An
701 experimental study based on stable isotopes. Mar Environ Res 142:208-213.
702 12. Altamia MA, Wood N, Fung JM, Dedrick S, Linton EW, Concepcion GP, Haygood MG,
703 Distel DL. 2014. Genetic differentiation among isolates of Teredinibacter turnerae, a
704 widely occurring intracellular endosymbiont of shipworms. Mol Ecol 23:1418-32.
705 13. Distel DL, Morrill W, MacLaren-Toussaint N, Franks D, Waterbury J. 2002. Teredinibacter
706 turnerae gen. nov., sp. nov., a dinitrogen-fixing, cellulolytic, endosymbiotic gamma-
707 proteobacterium isolated from the gills of wood-boring molluscs (Bivalvia: Teredinidae).
708 Int J Syst Evol Microbiol 52:2261-9.
709 14. Yang JC, Madupu R, Durkin AS, Ekborg NA, Pedamallu CS, Hostetler JB, Radune D, Toms
710 BS, Henrissat B, Coutinho PM, Schwarz S, Field L, Trindade-Silva AE, Soares CA,
711 Elshahawi S, Hanora A, Schmidt EW, Haygood MG, Posfai J, Benner J, Madinger C, Nove
712 J, Anton B, Chaudhary K, Foster J, Holman A, Kumar S, Lessard PA, Luyten YA, Slatko B,
713 Wood N, Wu B, Teplitski M, Mougous JD, Ward N, Eisen JA, Badger JH, Distel DL. 2009.
714 The complete genome of Teredinibacter turnerae T7901: an intracellular endosymbiont
715 of marine wood-boring bivalves (shipworms). PLoS One 4:e6085.
716 15. Trindade-Silva AE, Machado-Ferreira E, Senra MV, Vizzoni VF, Yparraguirre LA, Leoncini
717 O, Soares CA. 2009. Physiological traits of the symbiotic bacterium Teredinibacter
718 turnerae isolated from the mangrove shipworm Neoteredo reynei. Genet Mol Biol
719 32:572-81.

720 16. Cimermancic P, Medema MH, Claesen J, Kurita K, Wieland Brown LC, Mavrommatis K,
721 Pati A, Godfrey PA, Koehrsen M, Clardy J, Birren BW, Takano E, Sali A, Linington RG,
722 Fischbach MA. 2014. Insights into secondary metabolism from a global analysis of
723 prokaryotic biosynthetic gene clusters. Cell 158:412-421.
724 17. Han AW, Sandy M, Fishman B, Trindade-Silva AE, Soares CA, Distel DL, Butler A, Haygood
725 MG. 2013. Turnerbactin, a novel triscatecholate siderophore from the shipworm
726 endosymbiont Teredinibacter turnerae T7901. PLoS One 8:e76151.
727 18. Elshahawi SI, Trindade-Silva AE, Hanora A, Han AW, Flores MS, Vizzoni V, Schrago CG,
728 Soares CA, Concepcion GP, Distel DL, Schmidt EW, Haygood MG. 2013. Boronated
729 tartrolon antibiotic produced by symbiotic cellulose-degrading bacteria in shipworm
730 gills. Proc Natl Acad Sci U S A 110:E295-304.
731 19. Voight JRR. 2015. Xylotrophic bivalves: aspects of their biology and the impacts of
732 humans. J Molluscan Stud 81:175-186.
733 20. Lopes SGBC, Domanseschi O, de Moraes DT, Morita M, Meserani GDLC. 2000. Functional
734 anatomy of the digestive system of Neoteredo reynei (Bartsch, 1920) and Psiloteredo
735 healdi (Bartsch, 1931) (Bivalvia: Teredinidae), p 257-271. *In* Harper EM, Taylor JD, Crame
736 JA (ed), The Evolutionary Biology of the Bivalvia, vol 177. Geological Society, London.
737 21. Filho CS, Tagliaro CH, Beasley CR. 2008. Seasonal abundance of the shipworm
738 Neoteredo reynei (Bivalvia, Teredinidae) in mangrove driftwood from a northern
739 Brazilian beach. Iheringia Série Zoologia 98:17-23.
740 22. Altamia MA, Shipway JR, Concepcion GP, Haygood MG, Distel DL. 2019. Thiosocius
741 teredinicola gen. nov., sp. nov., a sulfur-oxidizing chemolithoautotrophic endosymbiont
742 cultivated from the gills of the giant shipworm, Kuphus polythalamius. Int J Syst Evol
743 Microbiol 69:638-644.
744 23. Shipway JR, Altamia MA, Rosenberg G, Concepcion GP, Haygood MG, Distel DL. 2019. A
745 rock-boring and rock-ingesting freshwater bivalve (shipworm) from the Philippines. Proc
746 Biol Sci 286:20190434.
747 24. Shipway JR, O'Connor R, Stein D, Cragg SM, Korshunova T, Martynov A, Haga T, Distel
748 DL. 2016. Zachsia zenkewitschi (Teredinidae), a Rare and Unusual Seagrass Boring
749 Bivalve Revisited and Redescribed. PLoS One 11:e0155269.
750 25. Elshahawi SI. 2012. Isolation and biosynthesis of bioactive natural products produced by
751 marine symbionts. PhD. Oregon Health & Science University, Portland.
752 26. Brito TL, Campos AB, Bastiaan von Meijenfeldt FA, Daniel JP, Ribeiro GB, Silva GGZ,
753 Wilke DV, de Moraes DT, Dutilh BE, Meirelles PM, Trindade-Silva AE. 2018. The gill-
754 associated microbiome is the main source of wood plant polysaccharide hydrolases and
755 secondary metabolite gene clusters in the mangrove shipworm Neoteredo reynei. PLoS
756 One 13:e0200437.
757 27. Altamia MA, Shipway JR, Betcher MA, Stein D, Fung JM, Jospin G, Eisen JA, Haygood MG,
758 Distel DL. 2020. *Teredinibacter waterburyi* sp. nov., a marine, cellulolytic endosymbiotic
759 bacterium isolated from the gills of the wood-boring mollusc *Bankia setacea* (Bivalvia:
760 Teredinidae), and emended description of the genus Teredinibacter. Int J System Evol
761 Microbiol accepted.
762 28. Paul B, Dixit G, Murali TS, Satyamoorthy K. 2019. Genome-based taxonomic
763 classification. Genome 62:45-52.

764 29. Kwan JC, Donia MS, Han AW, Hirose E, Haygood MG, Schmidt EW. 2012. Genome
765 streamlining and chemical defense in a coral reef symbiosis. Proc Natl Acad Sci U S A
766 109:20655-60.
767 30. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de Los
768 Santos ELC, Kim HU, Nave M, Dickschat JS, Mitchell DA, Shelest E, Breitling R, Takano E,
769 Lee SY, Weber T, Medema MH. 2017. antiSMASH 4.0-improvements in chemistry
770 prediction and gene cluster boundary identification. Nucleic Acids Res 45:W36-W41.
771 31. Medema MH, Takano E, Breitling R. 2013. Detecting sequence homology at the gene
772 cluster level with MultiGeneBlast. Mol Biol Evol 30:1218-23.
773 32. Adamek M, Spohn M, Stegmann E, Ziemert N. 2017. Mining Bacterial Genomes for
774 Secondary Metabolite Gene Clusters. Methods Mol Biol 1520:23-47.
775 33. Kinscherf TG, Willis DK. 2005. The biosynthetic gene cluster for the beta-lactam
776 antibiotic tabtoxin in Pseudomonas syringae. J Antibiot (Tokyo) 58:817-21.
777 34. Kinscherf TG, Coleman RH, Barta TM, Willis DK. 1991. Cloning and expression of the
778 tabtoxin biosynthetic region from Pseudomonas syringae. J Bacteriol 173:4124-32.
779 35. Sinden SL, Durbin RD. 1968. Glutamine synthetase inhibition: possible mode of action of
780 wildfire toxin from Pseudomonas tabaci. Nature 219:379-80.
781 36. Turner JG, Debbage JM. 1982. Tabtoxin-induced symptoms are associated with the
782 accumulation of ammonia formed during photorespiration. Physiol Plant Pathol 20:223-
783 233.
784 37. Sudek S, Lopanik NB, Waggoner LE, Hildebrand M, Anderson C, Liu H, Patel A, Sherman
785 DH, Haygood MG. 2007. Identification of the putative bryostatin polyketide synthase
786 gene cluster from "Candidatus Endobugula sertula", the uncultivated microbial symbiont
787 of the marine bryozoan Bugula neritina. J Nat Prod 70:67-74.
788 38. Schoner TA, Gassel S, Osawa A, Tobias NJ, Okuno Y, Sakakibara Y, Shindo K, Sandmann
789 G, Bode HB. 2016. Aryl Polyenes, a Highly Abundant Class of Bacterial Natural Products,
790 Are Functionally Related to Antioxidative Carotenoids. Chembiochem 17:247-53.
791 39. Graf J, Ruby EG. 2000. Novel effects of a transposon insertion in the Vibrio fischeri glnD
792 gene: defects in iron uptake and symbiotic persistence in addition to nitrogen utilization.
793 Mol Microbiol 37:168-79.
794 40. Holden VI, Bachman MA. 2015. Diverging roles of bacterial siderophores during
795 infection. Metallomics 7:986-95.
796 41. Irschik H, Schummer D, Gerth K, Hofle G, Reichenbach H. 1995. The tartrolons, new
797 boron-containing antibiotics from a myxobacterium, Sorangium cellulosum. J Antibiot
798 (Tokyo) 48:26-30.
799 42. O'Connor R, Schmidt EW. 2018. Methods and compositions for prevention and
800 treatment of apicomplexan infections patent WO2018106966A1.
801 43. Popham JD, Dickson MR. 1973. Bacterial associations in the teredo Bankia australis
802 (Lamellibranchia: Mollusca). Mar Biol 19:338-340.
803 44. Schmidt EW. 2008. Trading molecules and tracking targets in symbiotic interactions. Nat
804 Chem Biol 4:466-73.
805 45. Donia MS, Cimermancic P, Schulze CJ, Wieland Brown LC, Martin J, Mitreva M, Clardy J,
806 Linington RG, Fischbach MA. 2014. A systematic analysis of biosynthetic gene clusters in
807 the human microbiome reveals a common family of antibiotics. Cell 158:1402-1414.

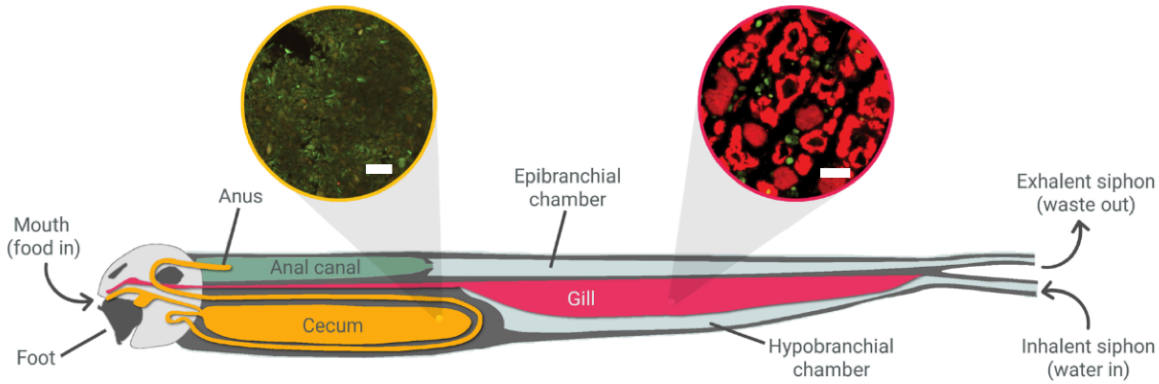
808 46. Costa-Lotufo LV, Carnevale-Neto F, Trindade-Silva AE, Silva RR, Silva GGZ, Wilke DV,
809 Pinto FCL, Sahm BDB, Jimenez PC, Mendonca JN, Lotufo TMC, Pessoa ODL, Lopes NP.
810 2018. Chemical profiling of two congeneric sea mat corals along the Brazilian coast:
811 adaptive and functional patterns. Chem Commun (Camb) 54:1952-1955.
812 47. Garcia GD, Gregoracci GB, Santos Ede O, Meirelles PM, Silva GG, Edwards R, Sawabe T,
813 Gotoh K, Nakamura S, Iida T, de Moura RL, Thompson FL. 2013. Metagenomic analysis of
814 healthy and white plague-affected Mussismilia braziliensis corals. Microb Ecol 65:1076-
815 86.
816 48. Joshi NA, Fass JN. 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool
817 for FASTQ files (Version 1.33) https://github.com/najoshi/sickle. Accessed
818 49. Peng Y, Leung Hc Fau - Yiu SM, Yiu Sm Fau - Chin FYL, Chin FY. 2012. IDBA-UD: a de novo
819 assembler for single-cell and metagenomic sequencing data with highly uneven depth.
820 Bioinformatics 28:1420-1428.
821 50. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM,
822 Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G,
823 Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its
824 applications to single-cell sequencing. J Comput Biol 19:455-77.
825 51. Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Prjibelski AD,
826 Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, Clingenpeel SR, Woyke T, Mclean JS,
827 Lasken R, Tesler G, Alekseyev MA, Pevzner PA. 2013. Assembling Single-Cell Genomes
828 and Mini-Metagenomes From Chimeric MDA Products. Journal of Computational
829 Biology 20:714-737.
830 52. Bushnell B, Rood J, Singer E. 2017. BBMerge - Accurate paired shotgun read merging via
831 overlap. PLoS One 12:e0185056.
832 53. Lo CC, Chain PS. 2014. Rapid evaluation and quality control of next generation
833 sequencing data with FaQCs. BMC Bioinformatics 15:366.
834 54. Wang Y, Leung H, Yiu S, Chin F. 2014. MetaCluster-TA: taxonomic annotation for
835 metagenomic data based on assembly-assisted binning. BMC Genomics 15 Suppl 1:S12.
836 55. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin
837 R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and
838 SAMtools. Bioinformatics 25:2078-9.
839 56. Li H. 2011. A statistical framework for SNP calling, mutation discovery, association
840 mapping and population genetical parameter estimation from sequencing data.
841 Bioinformatics 27:2987-93.
842 57. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal:
843 prokaryotic gene recognition and translation initiation site identification. BMC
844 Bioinformatics 11:119.
845 58. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using
846 DIAMOND. Nature Methods 12:59-60.
847 59. Tanizawa Y, Fujisawa T, Nakamura Y. 2018. DFAST: a flexible prokaryotic genome
848 annotation pipeline for faster genome publication. Bioinformatics 34:1037-1039.
849 60. Miller IJ, Rees ER, Ross J, Miller I, Baxa J, Lopera J, Kerby RL, Rey FE, Kwan JC. 2019.
850 Autometa: automated extraction of microbial genomes from individual shotgun
851 metagenomes. Nucleic Acids Res 47:e57.

852 61. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC,
853 Pati A. 2015. Microbial species delineation using whole genome sequences. Nucleic
854 Acids Research 43:6761-6771.
855 62. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify
856 genomes with the Genome Taxonomy Database. Bioinformatics
857 doi:10.1093/bioinformatics/btz848.
858 63. Lin Z, Kakule TB, Reilly CA, Beyhan S, Schmidt EW. 2019. Secondary Metabolites of
859 Onygenales Fungi Exemplified by Aioliomyces pyridodomos. J Nat Prod 82:1616-1626.
860 64. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B,
861 Ideker T. 2003. Cytoscape: a software environment for integrated models of
862 biomolecular interaction networks. Genome Res 13:2498-504.
863
864
865

866 **Figure Legends**
867
868 Figure 1. Top, diagram of generic shipworm anatomy. Insets are from Betcher *et al*., *PLoS One*,
869 2012 Figure 2, panels B and D, scale bar 20 μm (8). Red: signal from a fluorescent universal
870 bacterial probe indicating large numbers of bacterial symbionts in the bacteriocytes of the gill,
871 and paucity of bacteria in the cecum. Green is background fluorescence. Bottom, collection
872 locations of specimens included in this study. See Table S1 for details.
873
874 **Figure 2.** Cultivated bacterial isolates represent the major shipworm gill symbionts. **A)** Isolated
875 bacteria analyzed in this study are shown in abstracted schematic of a 16S rRNA phylogenetic
876 tree. The complete tree with accurate branch lengths and bootstrap numbers is shown in Fig.
877 S1. *T. turnerae* comprised 11 sequenced strains, for other groups individual strains are shown.
878 Each color indicates different bacteria appearing in the metagenomes in B. **B)** Species
879 composition of shipworm gill symbiont community based on shotgun metagenome sequence
880 analysis. The *y*-axis indicates the percent of reads originating from each bacterial species, while
881 the *x*-axis indicates individual shipworm specimens used in the study. Colors indicate the origin
882 of bacterial reads; gray is minor, sporadic, unidentified strains.
883
884 **Figure 3.** Heatmap of relationships between symbiont isolate genomes and gill metagenome
885 bins. The scale bar is shaded according to identity based upon (AF x gANI). Color bars in the
886 phylogenetic tree indicate bacterial species identity, either in the metagenomes or in the
887 genome, and they are identical to the codes shown in Fig. 2. This figure indicates the high
888 degree of certainty that the cultivated isolates are the same species as the major bacteria
889 present in the gill.
890
891 **Figure 4.** Most BGCs found in the metagenomes and in the bacterial isolate genomes are
892 shared. 401 BGCs from metagenome sequences were compared to the bacterial isolate
893 genomes, of which 305 could be found in isolates. Conversely, 148 of 168 BGCs from sequenced
894 bacterial isolates could be found in the metagenomes. The shared numbers likely differ because
895 the contigs assembled from the metagenome sequences were shorter on average, so that
896 several metagenome fragments may map to a single BGC in an isolate.
897
898 **Figure 5.** GCFs found in **A)** bacterial genomes and **B)** gill metagenomes. **A)** A list of strains of
899 cultivated bacterial genomes is provided in the *x*-axis, while the number of total GCFs in
900 different sequenced strains is shown in the *y*-axis. Colors indicate bacteria from Fig. 2A.
901 Because there are 11 isolates of a *T. turnerae*, the number of GCFs in this group (dark blue bars)
902 are comparatively overrepresented in the diagram. **B)** GCFs (*x*-axis) found in each metagenome
903 (*y*-axis) are shown. The inset expands a region containing the most common GCFs found in our
904 specimens. Colors indicate shipworm host species. See Table S4 for a complete list of GCFs used
905 in this figure.
906
907 **Figure 6.** A possible tabtoxin pathway is found in the *D. manni* metagenome. Tabtoxin is a
908 phytotoxin β-lactam initially discovered in *Pseudomonas* spp. (top). Strain 2719K contained a
909 tabtoxin-like cluster that was pseudogenized (shown as an insertion in *tabB*; middle). A non-

910 pseudogenized tabtoxin-like cluster was found in the *D. manni* metagenome gill (bottom)
911 supporting the observation that multiple variants of each symbiont genome are represented in
912 each metagenome.
913
914 **Figure 7.** GCF distribution across shipworm species. Shown is a similarity network diagram, in
915 which circles indicate individual BGCs from sequenced isolates (gray) and gill metagenomes
916 (colors indicate species of origin; see legend). Lines indicate the MultiGeneBlast scores between
917 identified BGCs, with thinner lines indicating a lower degree of similarity. For example, the
918 cluster labeled "GCF_8" encodes the pathway for the siderophore turnerbactin, the structure of
919 which is shown at right. The main cluster, circled by a light blue oval, includes BGCs that are
920 very similar to the originally described turnerbactin gene cluster. More distantly related BGCs,
921 with fewer lines connecting them to the majority nodes in GCF_8, might represent other
922 siderophores. GCF_11 likely all represent tartrolon D/E, a boronated polyketide shown at right.
923 For detailed alignments of BGCs, see Fig. S4.
924
925 **Figure 8.** Integration of tBLASTn and networking analyses reveals the pattern of occurrence of
926 GCFs in isolates and metagenomes. Here, we show only the most commonly occurring GCFs.
927 The values in each box indicate the BGC occurrence per specimen for each GCF (see Fig. S5 for
928 details). When the number equals 1, then the BGC is found in all specimens of that species.
929 When the number is less than one, then it indicates the fraction of specimens in which the
930 pathway is found. A number greater than one is specific to GCF_3, when two different types are
931 possible (see Fig. 8). In that case, in two *D. mannii* specimens and one *N. reynei* specimen, there
932 are two different classes of GCF_3, and only one class in the other specimens.
933
934 **Figure 9.** Three types of GCF_3 gene clusters are distributed in all cellulolytic shipworms in this
935 study. tBLASTx was used to compare the clusters, demonstrating the presence of three closely
936 related GCF_3 gene families found in all cellulolytic shipworm gills.
937
938 **Table 1. Example gANI values for shipworm gills in comparison to sequenced isolates,**
939 **extracted from Table S2.**
940
941
942

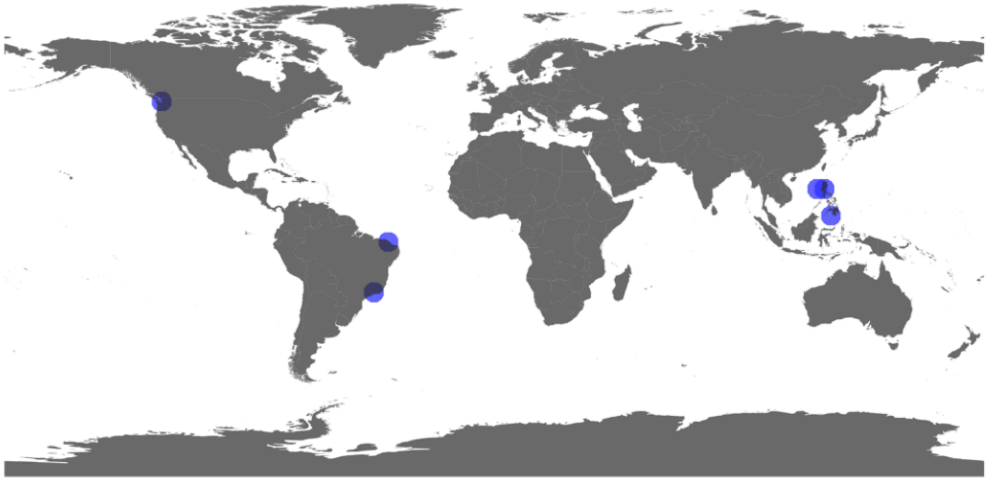
943 **Supporting Information.**
944
945
946 **Figure S1.** Phylogeny of shipworm gill symbionts and related free-living bacteria based on
947 approximate maximum-likelihood tree of 16S rRNA sequences. The tree was reconstructed
948 using 1,125 nucleotide positions employing GTR substitution model in FastTree version 2.1.11
949 with optimized Gamma20 likelihood and rate categories per site set to 20. Support values are
950 indicated for each node. The scale bar represents nucleotide substitution rate per site.
951 Cultivated shipworm symbionts and related bacteria are in boldface. The excerpted version of
952 this tree is shown in Fig. 2A.
953
954 **Figure S2.** AF x gANI comparison reveals species-level differences. A heatmap with AF x gANI
955 values comparing strain isolate genomes to each other. This analysis shows that *T. turnerae*
956 forms 2 distinct groups, which may possibly represent different species. However, the other
957 isolates are much more distantly related, with AF x gANI scores usually <0.2. Sulfide oxidizing
958 bacteria also bear some similarity.
959
960 **Figure S3.** Strain variation in shipworm gill symbiont bacterial species. This figure was made as
961 previously reported for *Kuphus* symbionts (7), using DNA gyrase B in 50 bp frames and
962 examining SNP variation. Different colors indicate reads with different SNPs along the gyrase
963 sequence. The *y*-axis represents number of reads observed, while the *x*-axis indicates each 50
964 bp region.
965
966 **Figure S4.** Representative alignments showing actual data underlying the clusters shown in Figs.
967 4, 5, 7, and 8. **A**) representative alignment of GCF_3 from genomes and metagenomes. Three
968 subtypes were indicated by red blue and green colors; for example, the NR03 metagenome
969 contains two copies of blue subtype. DM2858G and DM2722G contain blue and red subtypes.
970 **B**) alignment of GCF_2. **C**) alignment of GCF_5. **D**) alignment of GCF_8.
971
972 **Figure S5.** Occurrence of GCFs in individual samples, expanding what is shown in Fig. 8. **A)** GCFs
973 found in bacterial strains. **B)** GCFs from individual shipworm specimens.
974
975 **Figure S6.** Raw antiSMASH output showing total BGCs in shipworm isolates.
976
977 **Table S1A: Shipworm gill metagenomes used in this study.**
978
979 **Table S1B: Shipworm symbiont genomes.**
980
981 **Table S2: gANI comparison of genomes and metagenome bins.**
982
983 **Table S3. Comparison of contigs in the BSC2 bin.**
984
985 **Table S4. List of GCFs found in this study**.
986

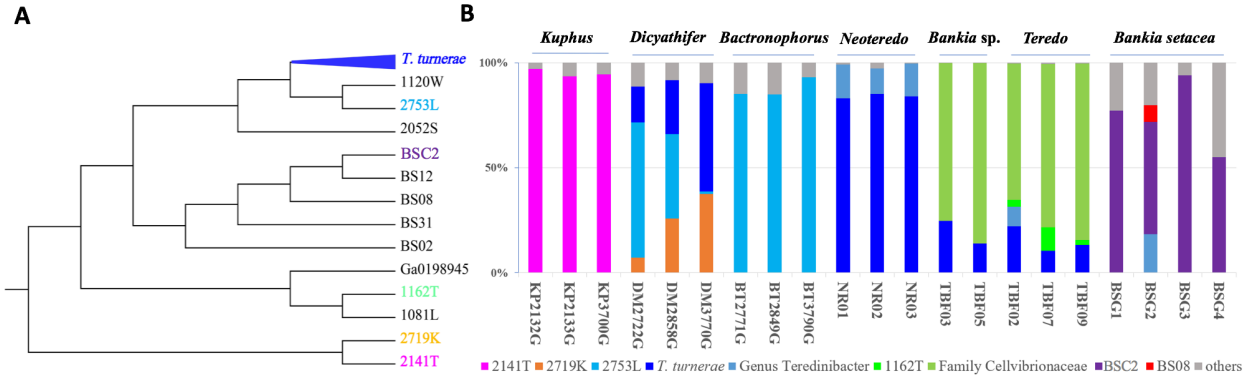


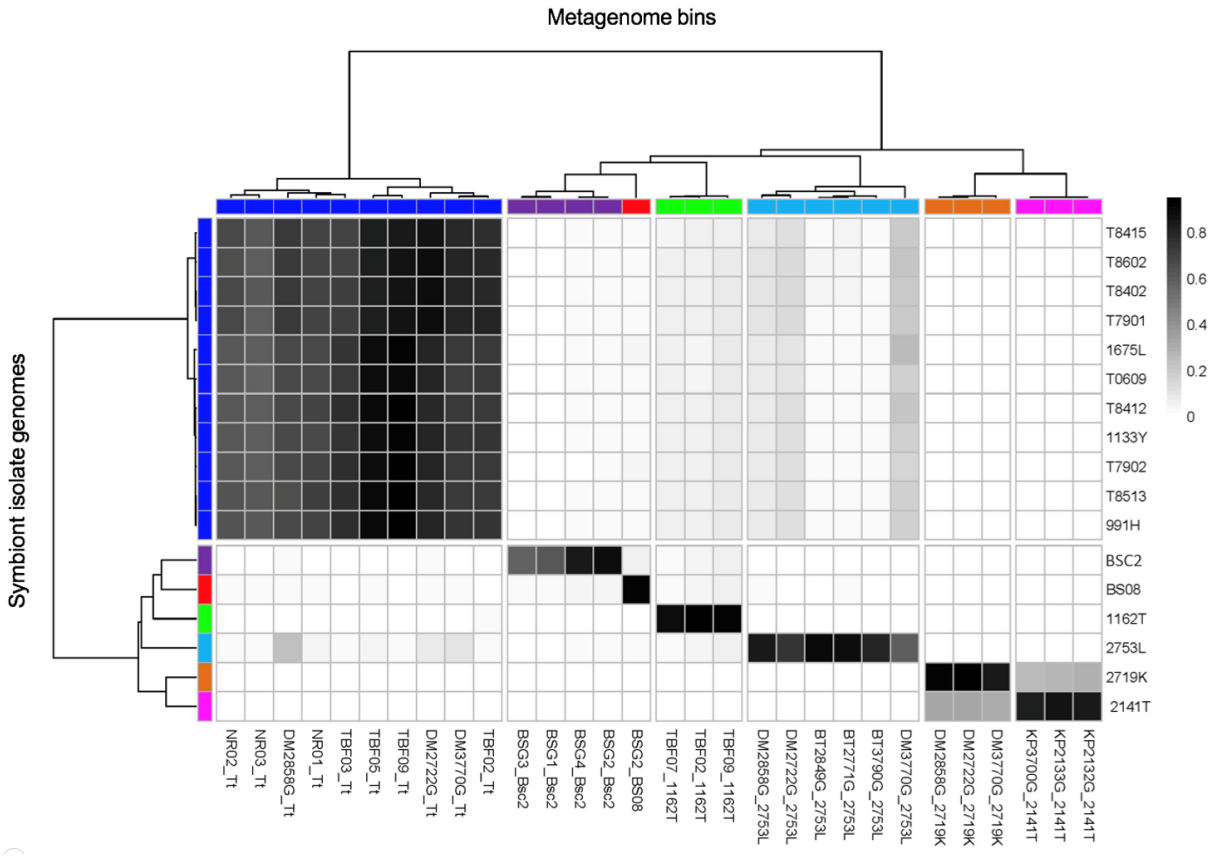
Wood-consuming shipworms
Cecum, vestigial in *Kuphus*,
wood-processing, quasi-sterile

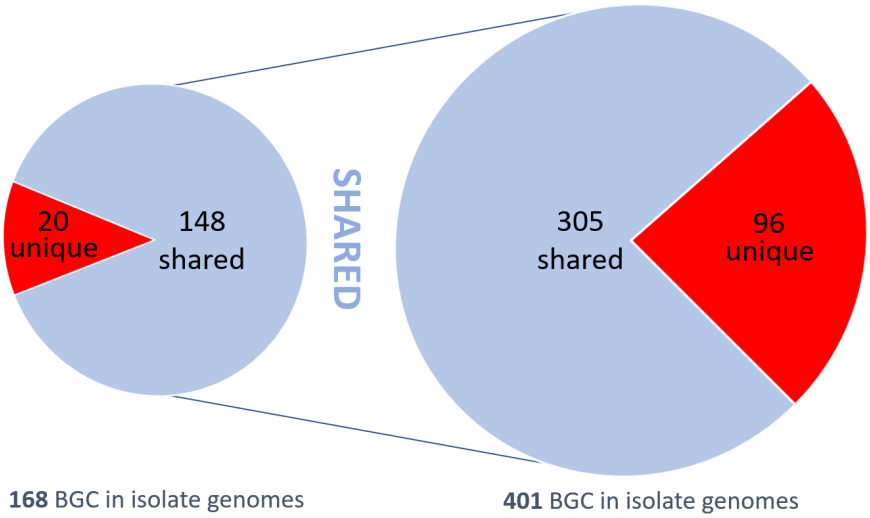
All shipworms
Gill with abundant bacteria

Mangrove specialists
Anal canal, secondary
waste processing

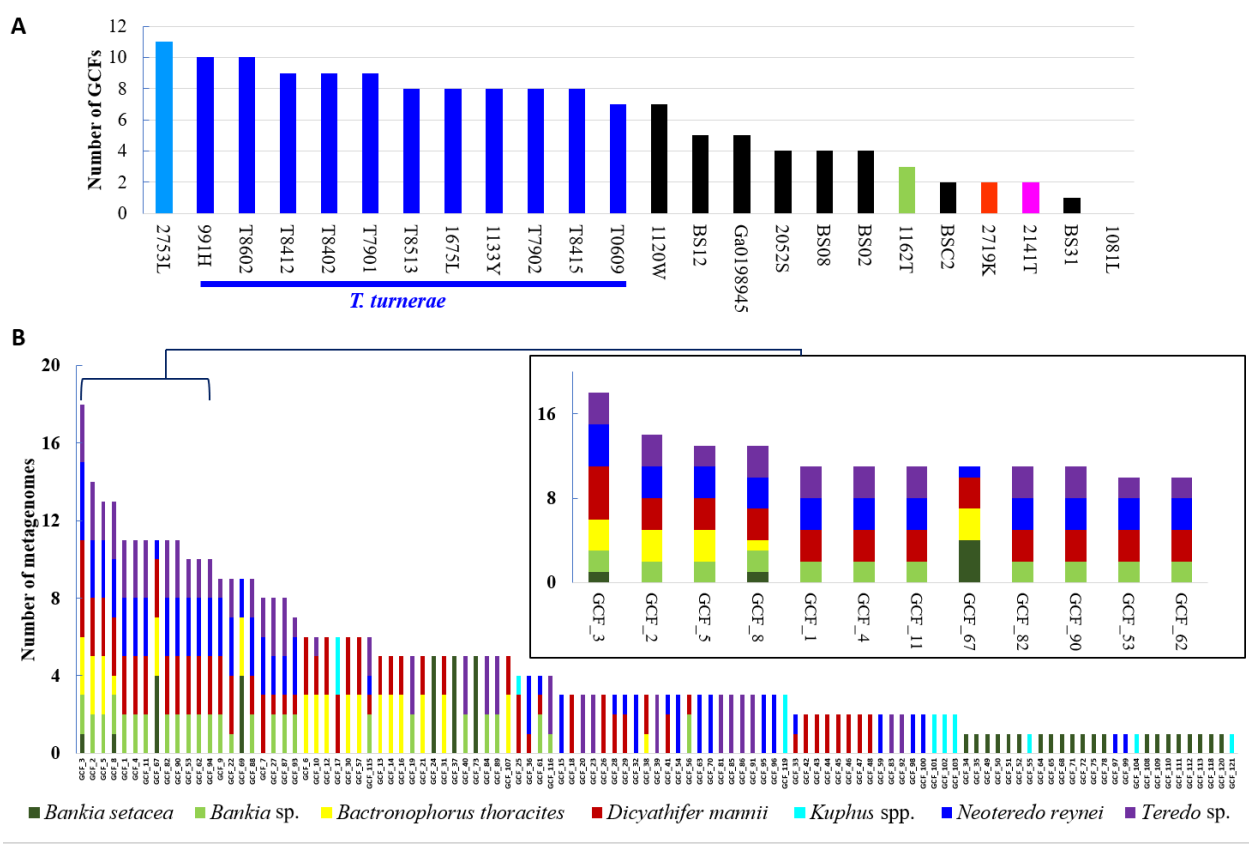


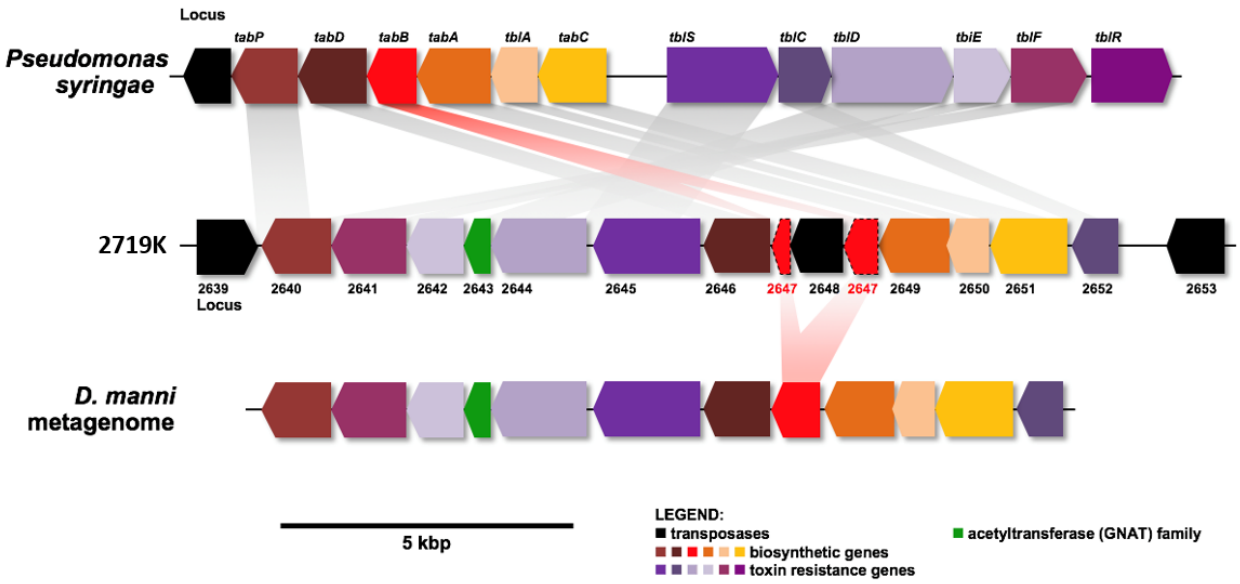

A
T. turnerae
1120W
2753L
2052S
BSC2
BS12
BS08
BS31
BS02
Ga0198945
1162T
1081L
2719K
2141T
B
Kuphus
Dicyathifer
Bactronophorus
Neoteredo
Bankia sp.
Teredo
Bankia setacea
100%
50%
0%
KP2132G
KP2133G
KP3700G
DM2722G
DM2858G
DM3770G
BT2771G
BT2849G
BT3790G
NR01
NR02
NR03
TBF03
TBF05
TBF02
TBF07
TBF09
BSG1
BSG2
BSG3
BSG4
2141T
2719K
2753L
T. turnerae
Genus Teredinibacter
1162T
Family Cellvibrionaceae
BSC2
BS08
others

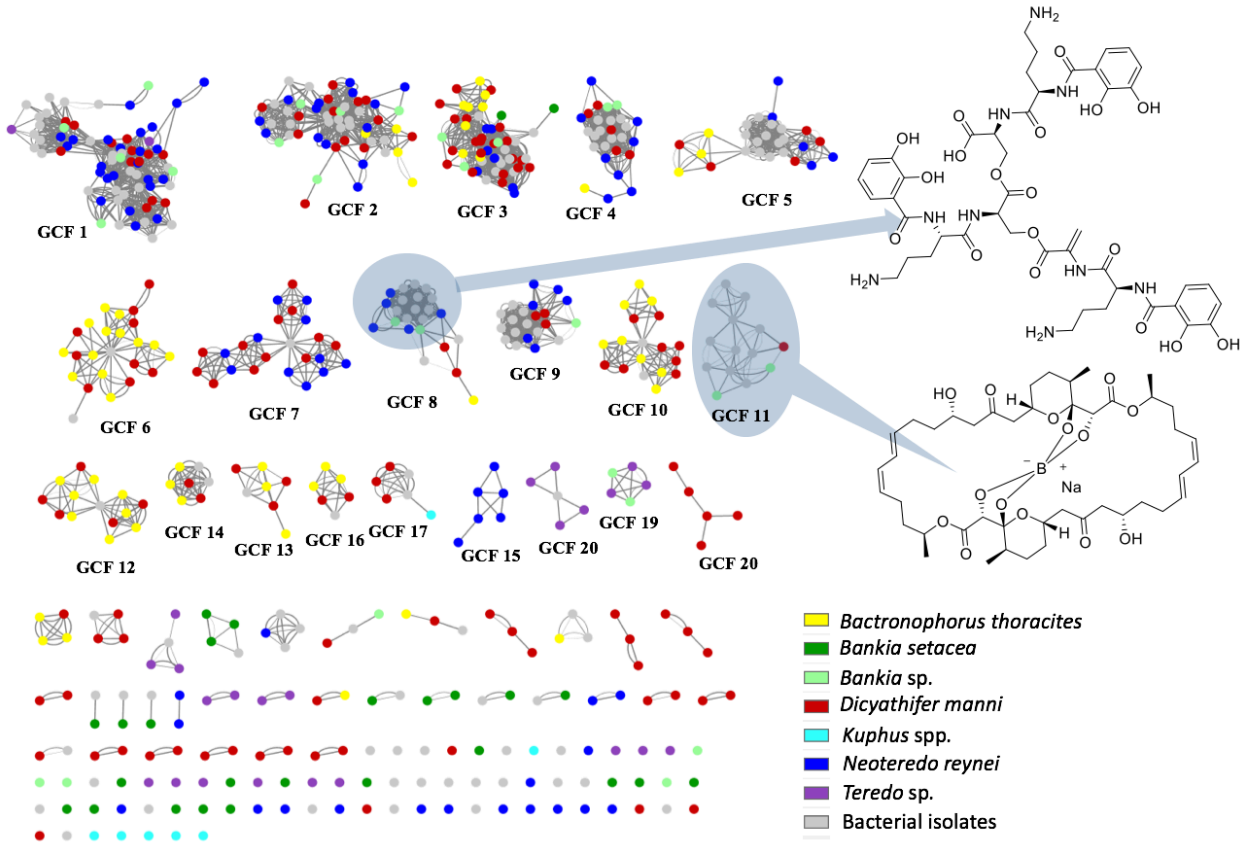

Metagenome bins
Symbiont isolate genomes
T8415
T8602
T8402
T7901
1675L
T0609
T8412
1133Y
T7902
T8513
991H
BSC2
BS08
1162T
2753L
2719K
2141T
0.8
0.6
0.4
0.2
0
NR02_Tt
NR03_Tt
DM2858G_Tt
NR01_Tt
TBF03_Tt
TBF05_Tt
TBF09_Tt
DM2722G_Tt
DM3770G_Tt
TBF02_Tt
BSG3_Bsc2
BSG1_Bsc2
BSG4_Bsc2
BSG2_Bsc2
BSG2_BS08
TBF07_1162T
TBF02_1162T
TBF09_1162T
DM2858G_2753L
DM2722G_2753L
BT2849G_2753L
BT2771G_2753L
BT3790G_2753L
DM3770G_2753L
DM2858G_2719K
DM2722G_2719K
DM3770G_2719K
KP3700G_2141T
KP2133G_2141T
KP2132G_2141T



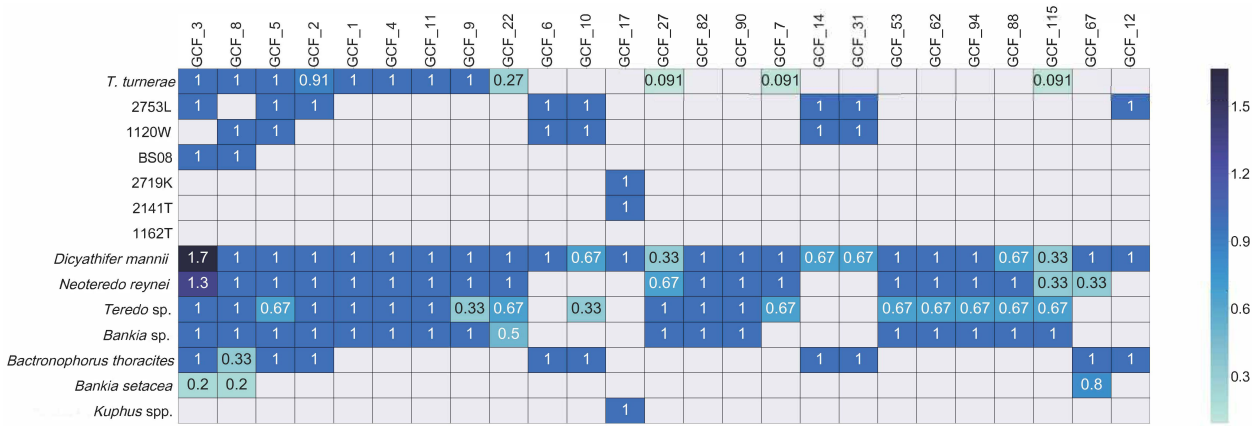
	unique	shared between isolate and metagenomes	total BGC counts
metagenome	96	305	401
isolate genome	20	148	168

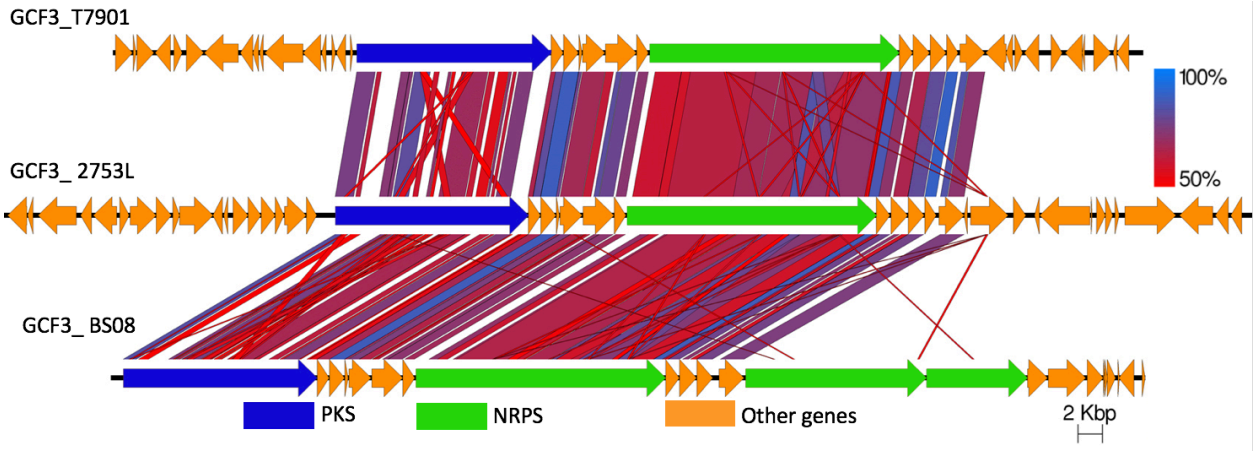

A
Number of GCFs
12
10
8
6
4
2
0
2753L
991H
T8602
T8412
T8402
T7901
T8513
1675L
1133Y
T7902
T8415
T0609
1120W
BS12
Ga0198945
2052S
BS08
BS02
1162T
BSC2
2719K
2141T
BS31
1081L
T. turnerae
B
Number of metagenomes
20
16
12
8
4
0
GCF_3
GCF_2
GCF_5
GCF_8
GCF_1
GCF_4
GCF_11
GCF_67
GCF_82
GCF_90
GCF_53
GCF_62
Bankia setacea
Bankia sp.
Bactronophorus thoracites
Dicyathifer mannii
Kuphus spp.
Neoteredo reynei
Teredo sp.


Locus
tabP
tabD
tabB
tabA
tblA
tabC
tblS
tblC
tblD
tbiE
tblF
tblR
Pseudomonas syringae
2719K
2639
Locus
2640
2641
2642
2643
2644
2645
2646
2647
2648
2647
2649
2650
2651
2652
2653
D. manni metagenome
5 kbp
LEGEND:
transposases
biosynthetic genes
toxin resistance genes
acetyltransferase (GNAT) family

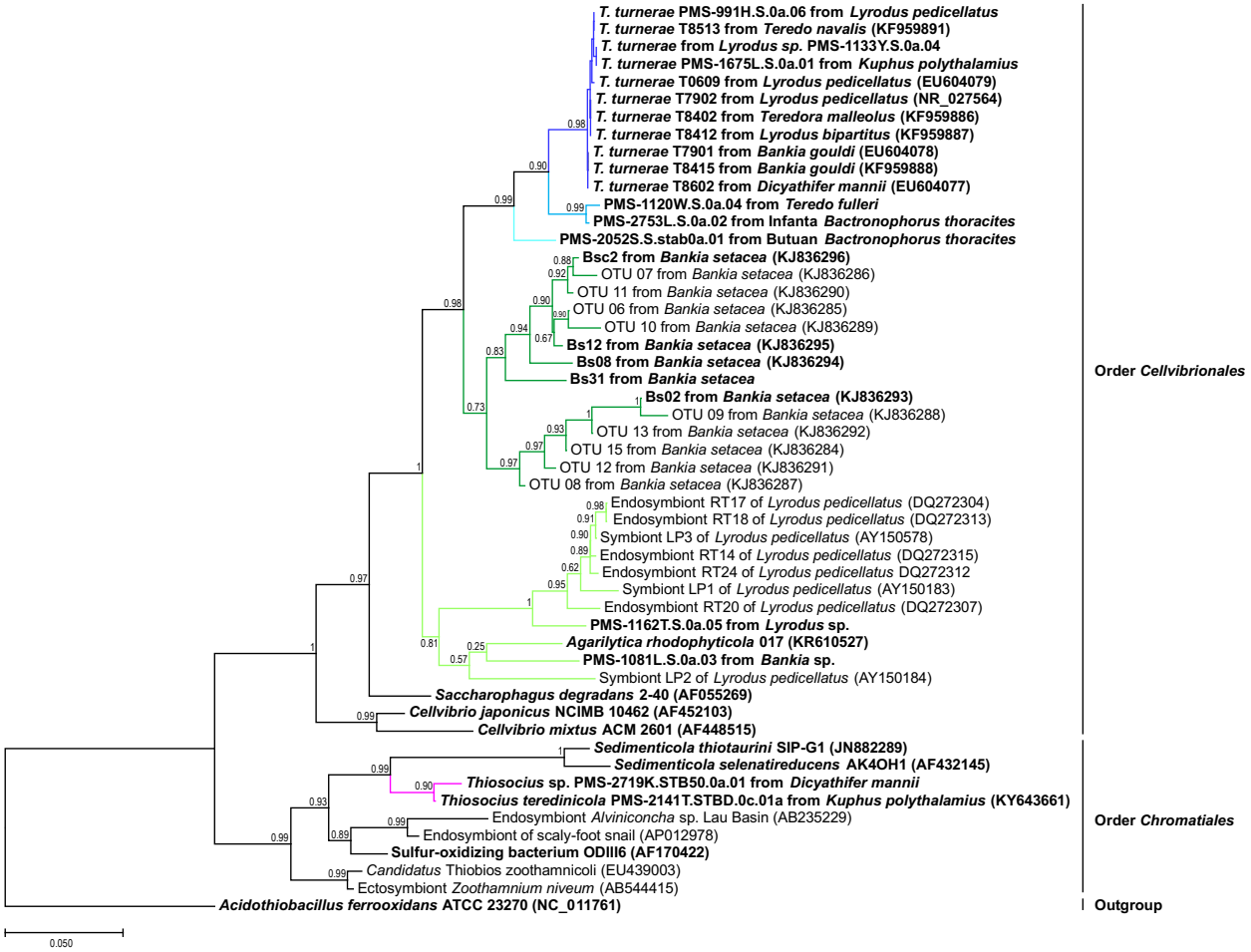

GCF 1
GCF 2
GCF 3
GCF 4
GCF 5
GCF 6
GCF 7
GCF 8
GCF 9
GCF 10
GCF 11
GCF 12
GCF 13
GCF 14
GCF 15
GCF 16
GCF 17
GCF 19
GCF 20
GCF 20
Na

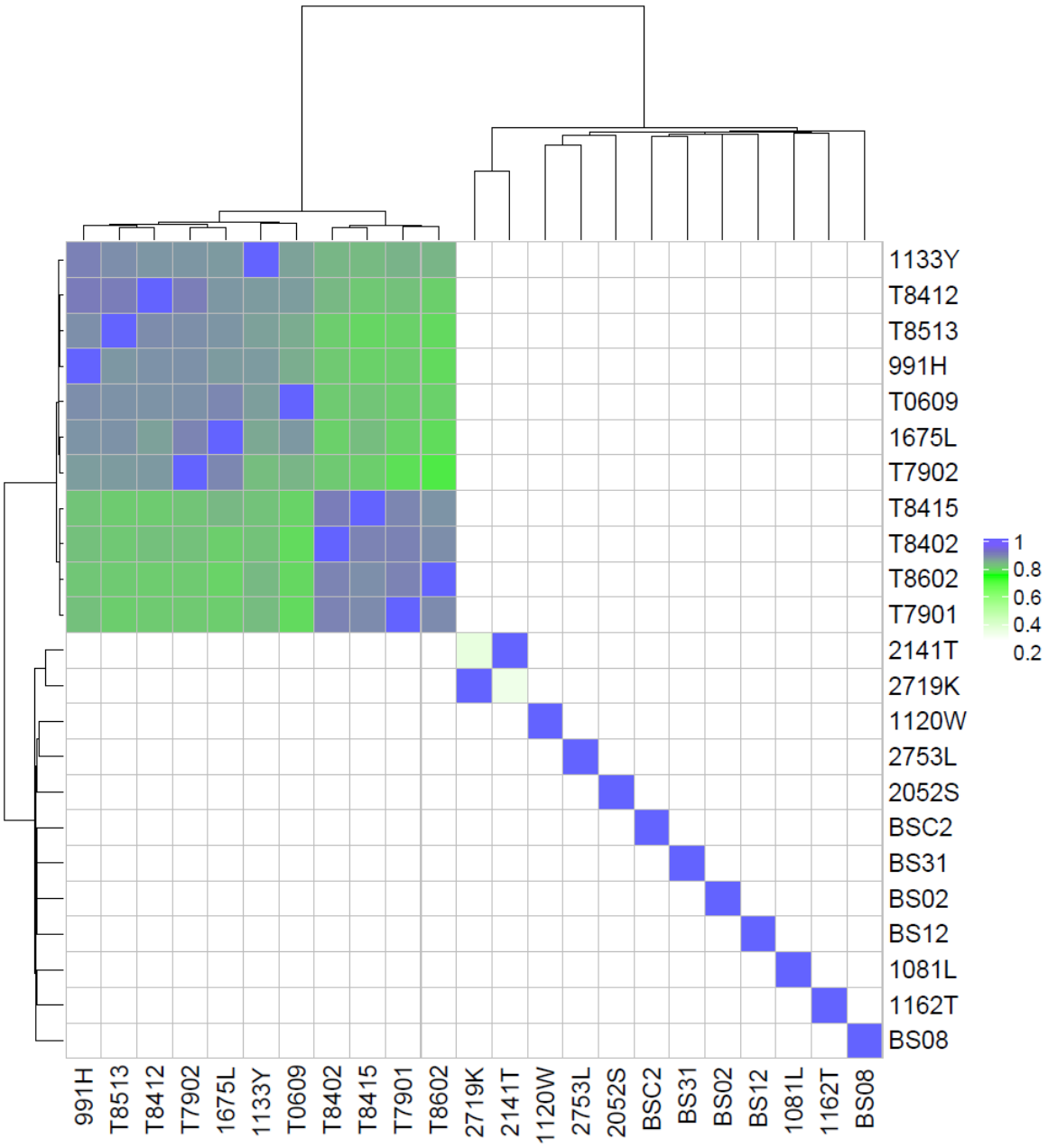
Bactronophorus thoracites
Bankia setacea
Bankia sp.
Dicyathifer manni
Kuphus spp.
Neoteredo reynei
Teredo sp.
Bacterial isolates

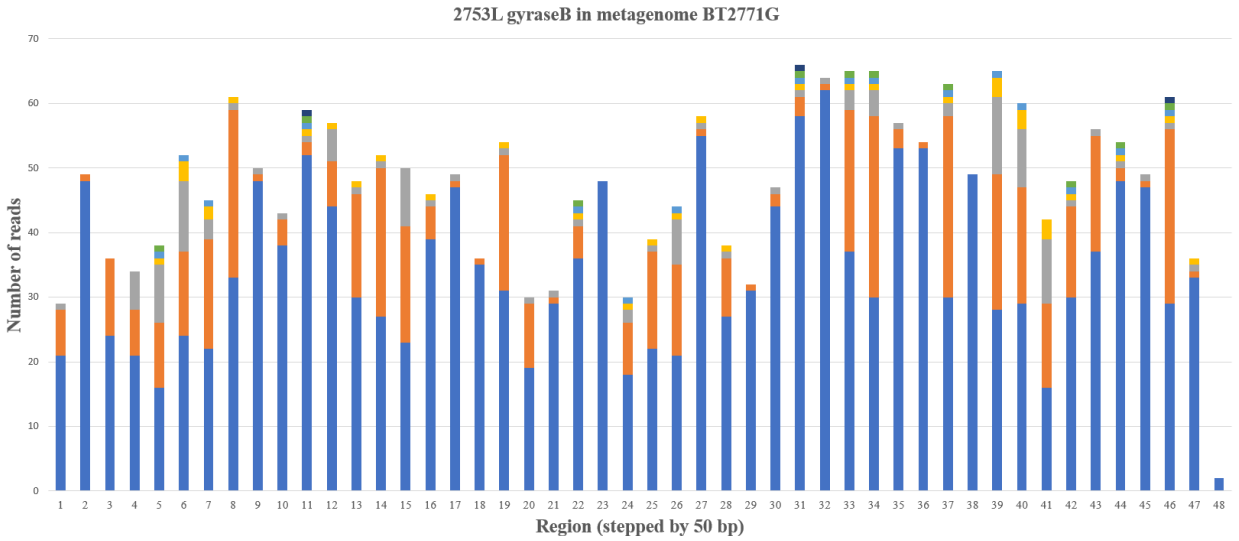

GCF_3
GCF_8
GCF_5
GCF_2
GCF_1
GCF_4
GCF_11
GCF_9
GCF_22
GCF_6
GCF_10
GCF_17
GCF_27
GCF_82
GCF_90
GCF_7
GCF_14
GCF_31
GCF_53
GCF_62
GCF_94
GCF_88
GCF_115
GCF_67
GCF_12
T. turnerae
2753L
1120W
BS08
2719K
2141T
1162T
Dicyathifer mannii
Neoteredo reynei
Teredo sp.
Bankia sp.
Bactronophorus thoracites
Bankia setacea
Kuphus spp.
1.5
1.2
0.9
0.6
0.3


GCF3_T7901
GCF3_ 2753L
GCF3_ BS08
100%
50%
PKS
NRPS
Other genes
2 Kbp

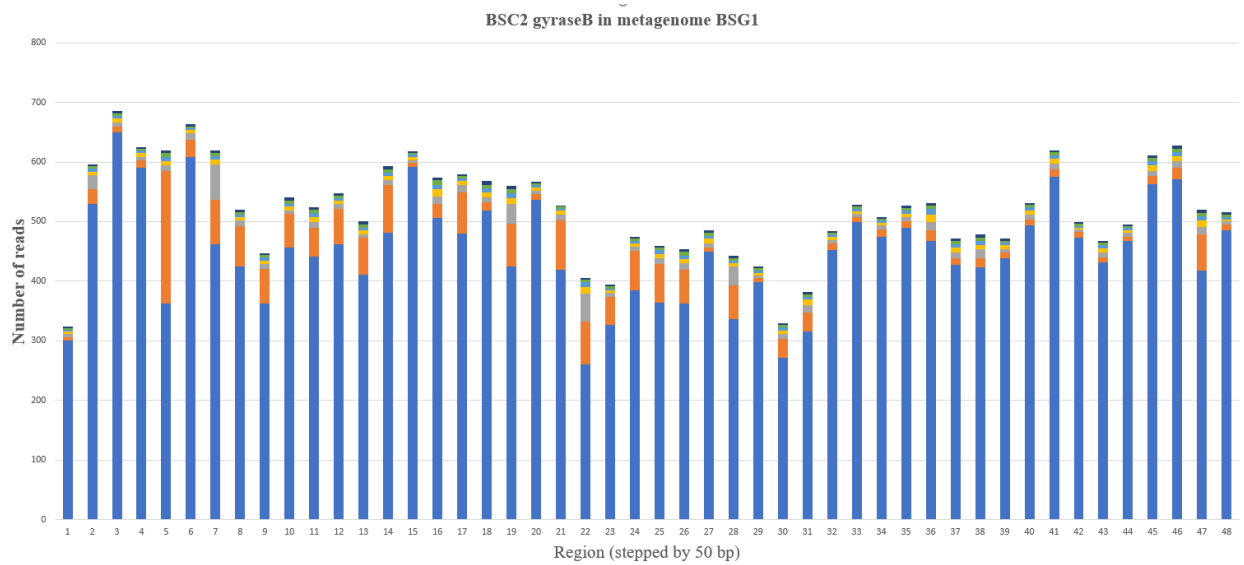
Comparison	**Total metagenomic bin size in bp**	**gANI**
Kuphus spp. : *T. teredicincola* 2141T	23758169	0.963839
D. mannii and *B. thoracites* : 2753L	26758239	0.990273
D. mannii : 2719K	14895102	0.992012
B. setacea : BSC2	19687153	0.974108
Teredo sp. : 1162T	1489154	0.984036
B. setacea : BS08	3513639	0.995137

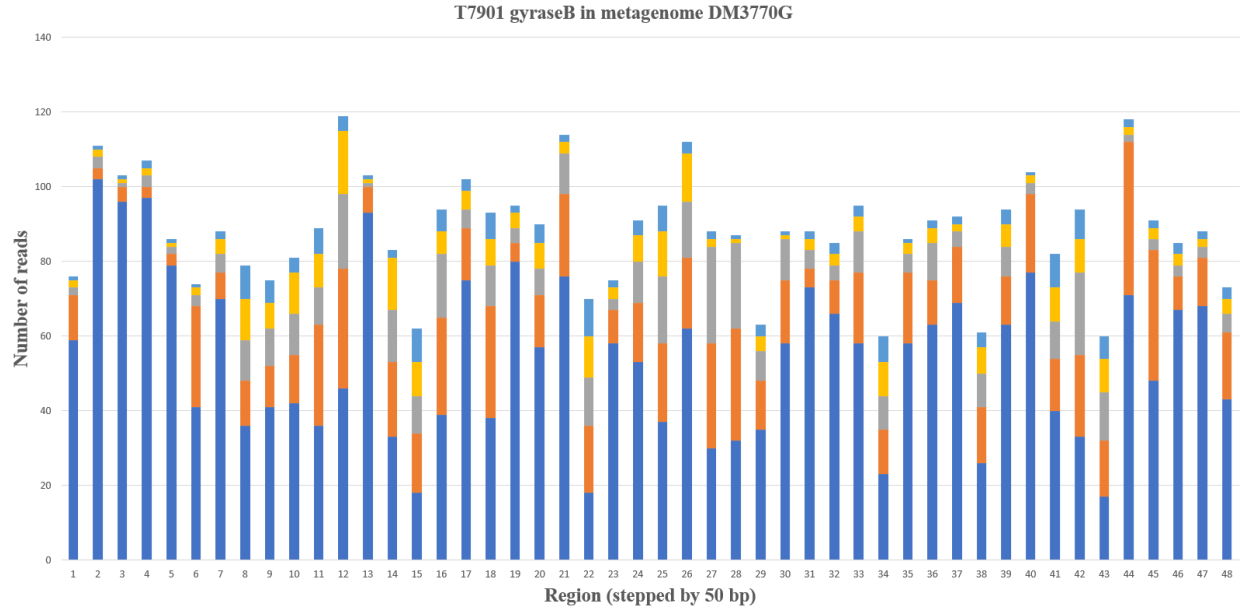

T. turnerae PMS-991H.S.0a.06 from Lyrodus pedicellatus
T. turnerae T8513 from Teredo navalis (KF959891)
T. turnerae from Lyrodus sp. PMS-1133Y.S.0a.04
T. turnerae PMS-1675L.S.0a.01 from Kuphus polythalamius
T. turnerae T0609 from Lyrodus pedicellatus (EU604079)
T. turnerae T7902 from Lyrodus pedicellatus (NR_027564)
T. turnerae T8402 from Teredora malleolus (KF959886)
T. turnerae T8412 from Lyrodus bipartitus (KF959887)
T. turnerae T7901 from Bankia gouldi (EU604078)
T. turnerae T8415 from Bankia gouldi (KF959888)
T. turnerae T8602 from Dicyathifer mannii (EU604077)
PMS-1120W.S.0a.04 from Teredo fulleri
PMS-2753L.S.0a.02 from Infanta Bactronophorus thoracites
PMS-2052S.S.stab0a.01 from Butuan Bactronophorus thoracites
Bsc2 from Bankia setacea (KJ836296)
OTU 07 from Bankia setacea (KJ836286)
OTU 11 from Bankia setacea (KJ836290)
OTU 06 from Bankia setacea (KJ836285)
OTU 10 from Bankia setacea (KJ836289)
Bs12 from Bankia setacea (KJ836295)
Bs08 from Bankia setacea (KJ836294)
Bs31 from Bankia setacea
Bs02 from Bankia setacea (KJ836293)
OTU 09 from Bankia setacea (KJ836288)
OTU 13 from Bankia setacea (KJ836292)
OTU 15 from Bankia setacea (KJ836284)
OTU 12 from Bankia setacea (KJ836291)
OTU 08 from Bankia setacea (KJ836287)
Endosymbiont RT17 of Lyrodus pedicellatus (DQ272304)
Endosymbiont RT18 of Lyrodus pedicellatus (DQ272313)
Symbiont LP3 of Lyrodus pedicellatus (AY150578)
Endosymbiont RT14 of Lyrodus pedicellatus (DQ272315)
Endosymbiont RT24 of Lyrodus pedicellatus DQ272312
Symbiont LP1 of Lyrodus pedicellatus (AY150183)
Endosymbiont RT20 of Lyrodus pedicellatus (DQ272307)
PMS-1162T.S.0a.05 from Lyrodus sp.
Agarilytica rhodophyticola 017 (KR610527)
PMS-1081L.S.0a.03 from Bankia sp.
Symbiont LP2 of Lyrodus pedicellatus (AY150184)
Saccharophagus degradans 2-40 (AF055269)
Cellvibrio japonicus NCIMB 10462 (AF452103)
Cellvibrio mixtus ACM 2601 (AF448515)
Sedimenticola thiotaurini SIP-G1 (JN882289)
Sedimenticola selenatireducens AK4OH1 (AF432145)
Thiosocius sp. PMS-2719K.STB50.0a.01 from Dicyathifer mannii
Thiosocius teredinicola PMS-2141T.STBD.0c.01a from Kuphus polythalamius (KY643661)
Endosymbiont Alviniconcha sp. Lau Basin (AB235229)
Endosymbiont of scaly-foot snail (AP012978)
Sulfur-oxidizing bacterium ODIII6 (AF170422)
Candidatus Thiobios zoothamnicoli (EU439003)
Ectosymbiont Zoothamnium niveum (AB544415)
Acidothiobacillus ferrooxidans ATCC 23270 (NC_011761)
Order Cellvibrionales
Order Chromatiales
Outgroup
0.050

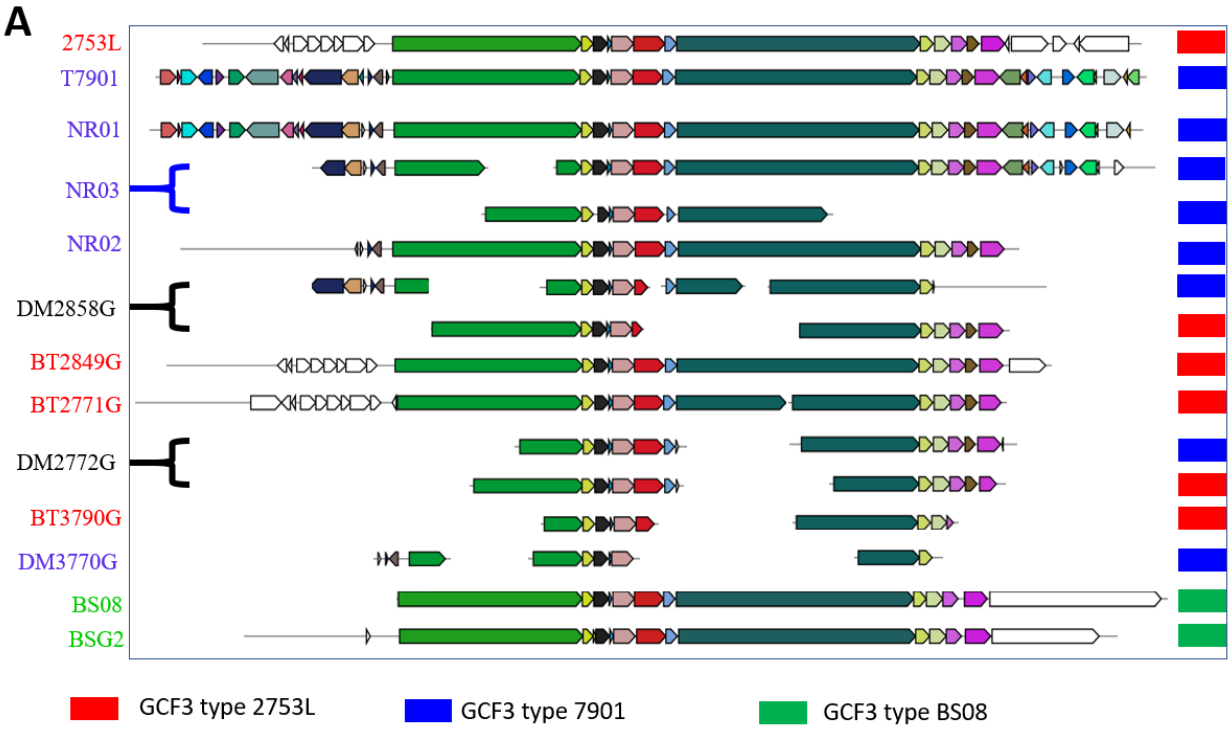

1133Y
T8412
T8513
991H
T0609
1675L
T7902
T8415
T8402
T8602
T7901
2141T
2719K
1120W
2753L
2052S
BSC2
BS31
BS02
BS12
1081L
1162T
BS08
991H
T8513
T8412
T7902
1675L
1133Y
T0609
T8402
T8415
T7901
T8602
2719K
2141T
1120W
2753L
2052S
BSC2
BS31
BS02
BS12
1081L
1162T
BS08
1
0.8
0.6
0.4
0.2

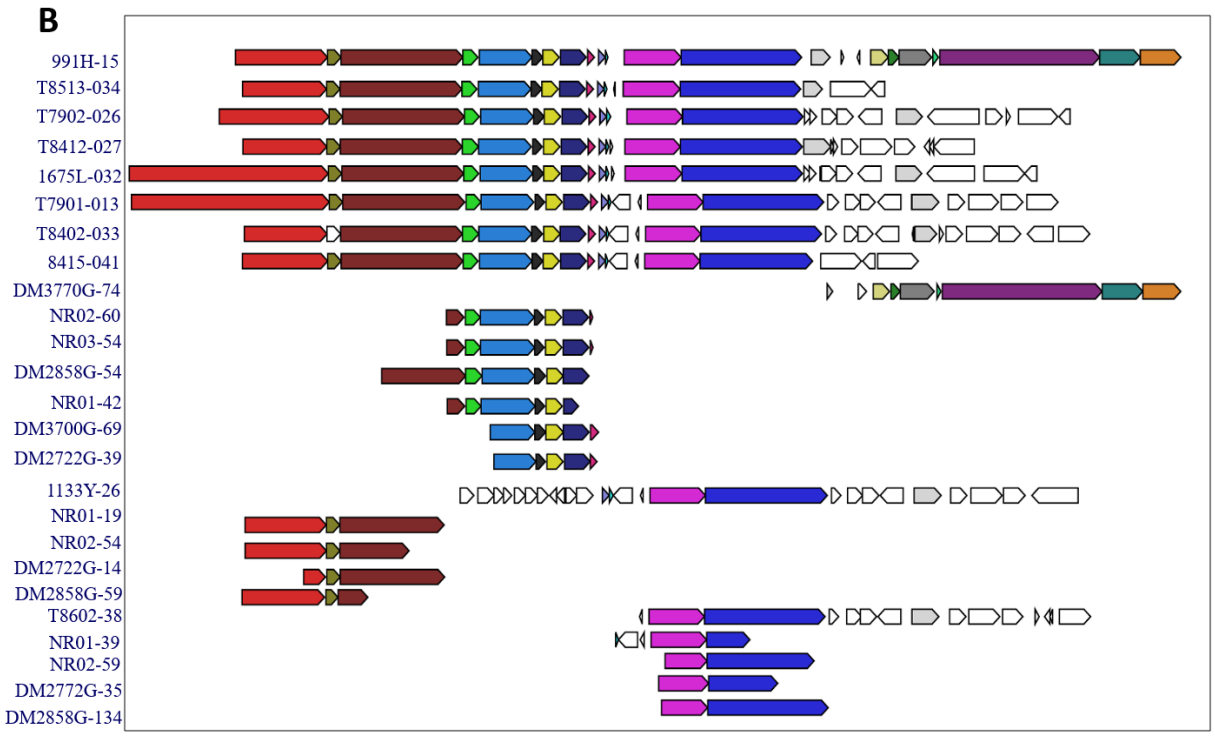

2753L gyraseB in metagenome BT2771G
Number of reads
Region (stepped by 50 bp)

2719K gyraseB in metagenome DM2858G
Number of reads
Region (stepped by 50 bp)

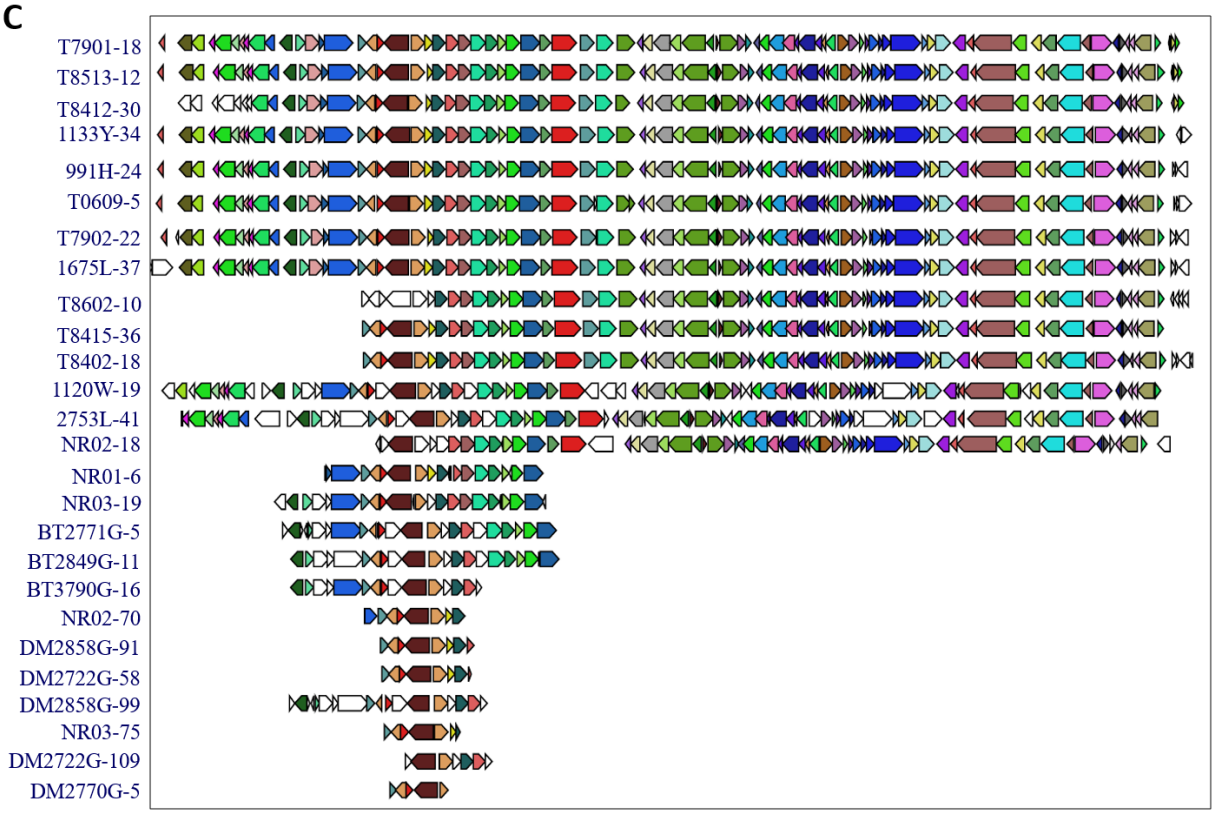

BSC2 gyraseB in metagenome BSG1
Number of reads
Region (stepped by 50 bp)


T7901 gyraseB in metagenome DM3770G
Number of reads
Region (stepped by 50 bp)

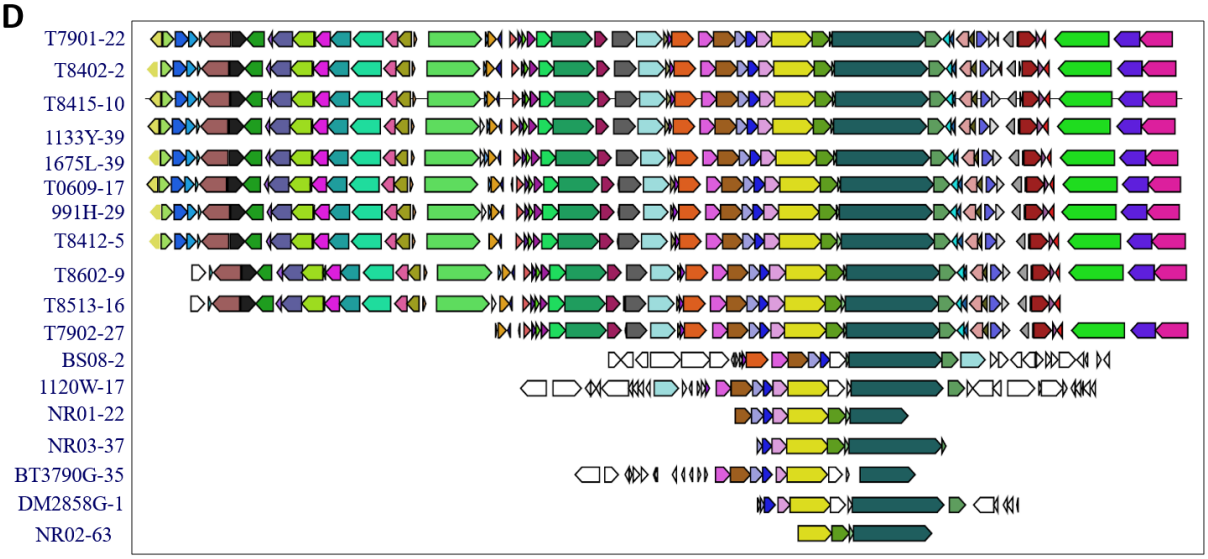

A
2753L
T7901
NR01
NR03
NR02
DM2858G
BT2849G
BT2771G
DM2772G
BT3790G
DM3770G
BS08
BSG2
GCF3 type 2753L
GCF3 type 7901
GCF3 type BS08


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

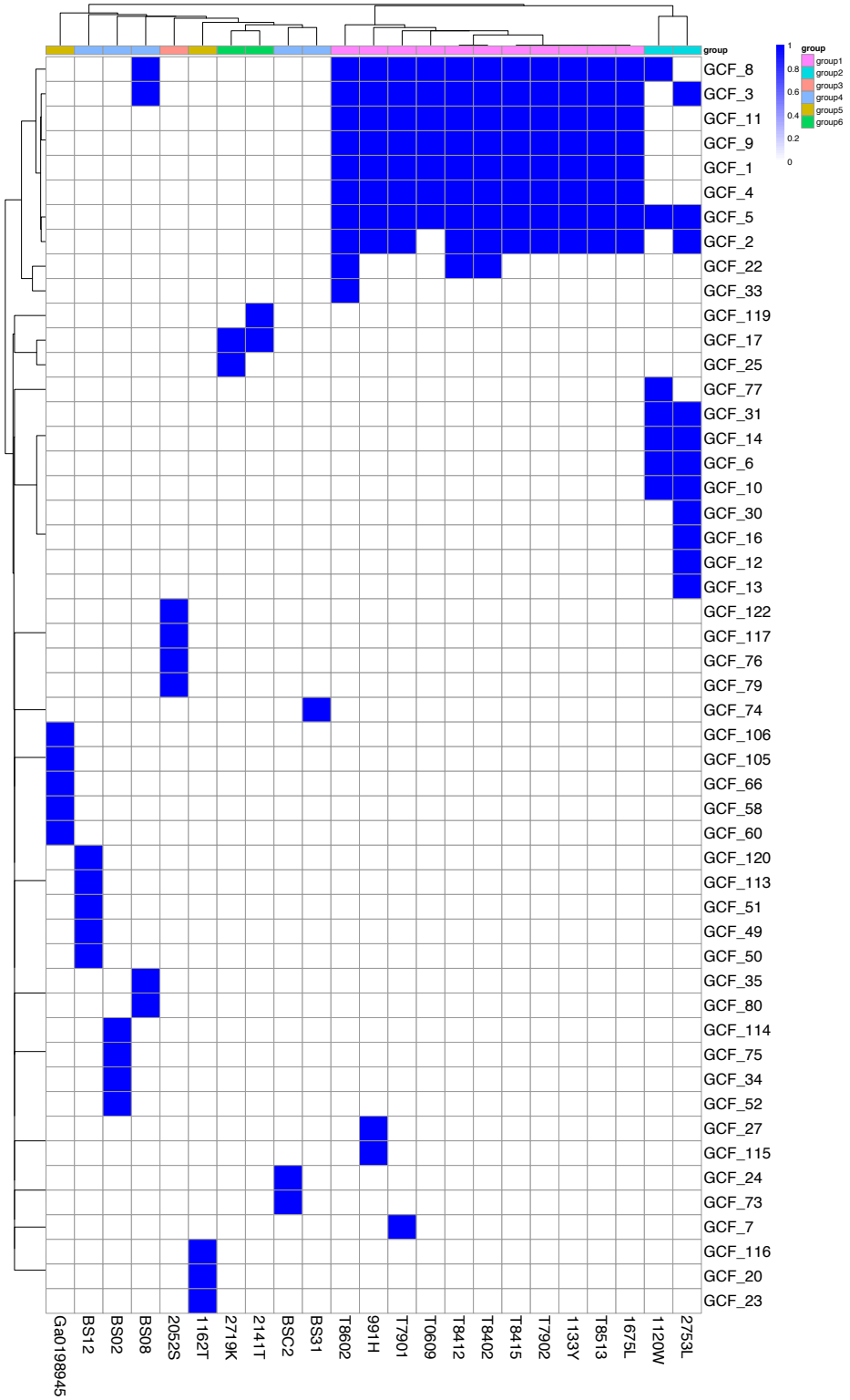
C


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

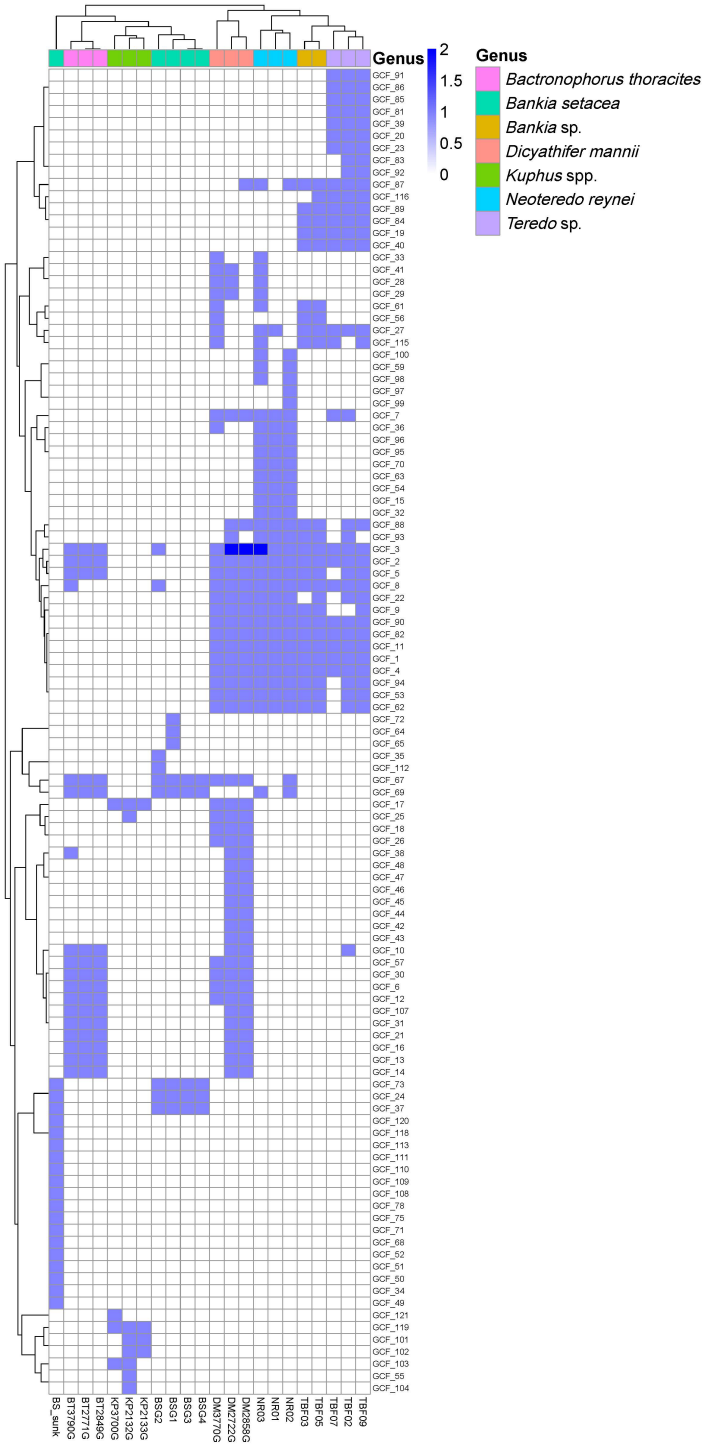
D

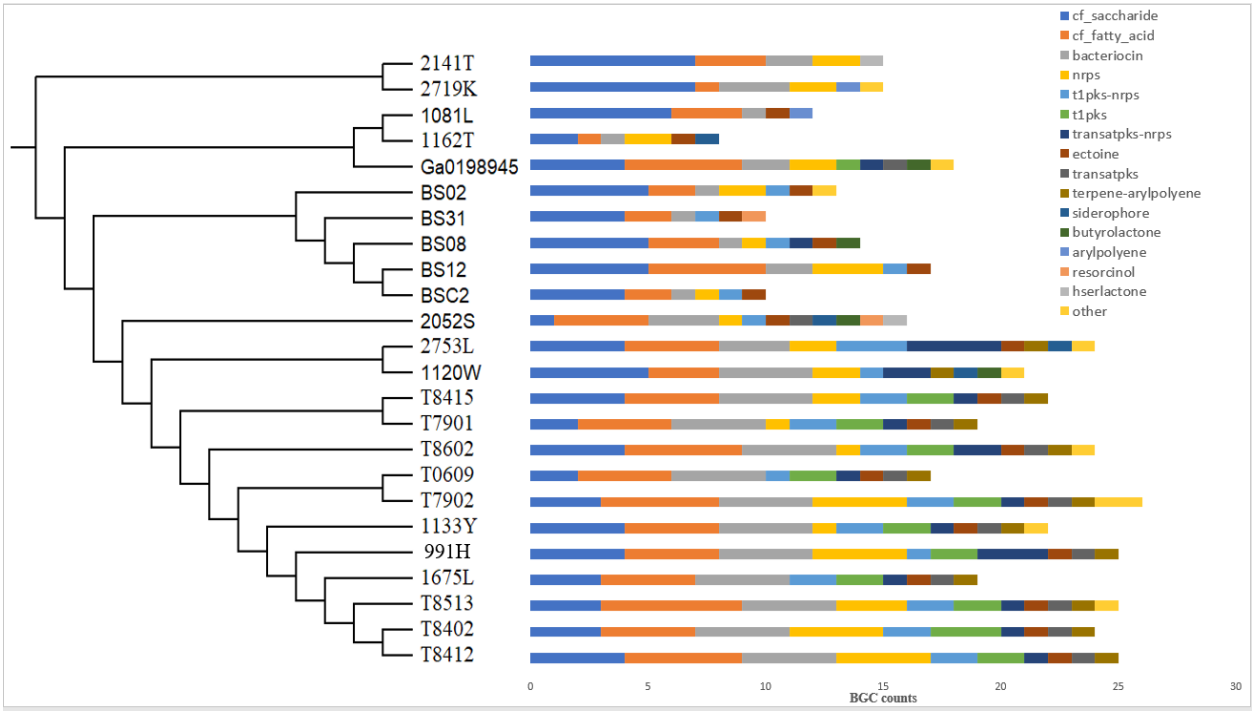

T7901-22
T8402-2
T8415-10
1133Y-39
1675L-39
T0609-17
991H-29
T8412-5
T8602-9
T8513-16
T7902-27
BS08-2
1120W-17
NR01-22
NR03-37
BT3790G-35
DM2858G-1
NR02-63

A


group
group1
group2
group3
group4
group5
group6
1
0.8
0.6
0.4
0.2
0
GCF_8
GCF_3
GCF_11
GCF_9
GCF_1
GCF_4
GCF_5
GCF_2
GCF_22
GCF_33
GCF_119
GCF_17
GCF_25
GCF_77
GCF_31
GCF_14
GCF_6
GCF_10
GCF_30
GCF_16
GCF_12
GCF_13
GCF_122
GCF_117
GCF_76
GCF_79
GCF_74
GCF_106
GCF_105
GCF_66
GCF_58
GCF_60
GCF_120
GCF_113
GCF_51
GCF_49
GCF_50
GCF_35
GCF_80
GCF_114
GCF_75
GCF_34
GCF_52
GCF_27
GCF_115
GCF_24
GCF_73
GCF_7
GCF_116
GCF_20
GCF_23
Ga0198945
BS12
BS02
BS08
2052S
1162T
2719K
2141T
BSC2
BS31
T8602
991H
T7901
T0609
T8412
T8402
T8415
T7902
1133Y
T8513
1675L
1120W
2753L

B


Genus
2
1.5
1
0.5
0
Genus
Bactronophorus thoracites
Bankia setacea
Bankia sp.
Dicyathifer mannii
Kuphus spp.
Neoteredo reynei
Teredo sp.
GCF_91
GCF_86
GCF_85
GCF_81
GCF_39
GCF_20
GCF_23
GCF_83
GCF_92
GCF_87
GCF_116
GCF_89
GCF_84
GCF_19
GCF_40
GCF_33
GCF_41
GCF_28
GCF_29
GCF_61
GCF_56
GCF_27
GCF_115
GCF_100
GCF_59
GCF_98
GCF_97
GCF_99
GCF_7
GCF_36
GCF_96
GCF_95
GCF_70
GCF_63
GCF_54
GCF_15
GCF_32
GCF_88
GCF_93
GCF_3
GCF_2
GCF_5
GCF_8
GCF_22
GCF_9
GCF_90
GCF_82
GCF_11
GCF_1
GCF_4
GCF_94
GCF_53
GCF_62
GCF_72
GCF_64
GCF_65
GCF_35
GCF_112
GCF_67
GCF_69
GCF_17
GCF_25
GCF_18
GCF_26
GCF_38
GCF_48
GCF_47
GCF_46
GCF_45
GCF_44
GCF_42
GCF_43
GCF_10
GCF_57
GCF_30
GCF_6
GCF_12
GCF_107
GCF_31
GCF_21
GCF_16
GCF_13
GCF_14
GCF_73
GCF_24
GCF_37
GCF_120
GCF_118
GCF_113
GCF_111
GCF_110
GCF_109
GCF_108
GCF_78
GCF_75
GCF_71
GCF_68
GCF_52
GCF_51
GCF_50
GCF_34
GCF_49
GCF_121
GCF_119
GCF_101
GCF_102
GCF_103
GCF_55
GCF_104
BS_sunk
BT3790G
BT2771G
BT2849G
KP3700G
KP2132G
KP2133G
BSG2
BSG1
BSG3
BSG4
DM3770G
DM2722G
DM2658G
NR03
NR01
NR02
TBF03
TBF05
TBF07
TBF02
TBF09


2141T
2719K
1081L
1162T
Ga0198945
BS02
BS31
BS08
BS12
BSC2
2052S
2753L
1120W
T8415
T7901
T8602
T0609
T7902
1133Y
991H
1675L
T8513
T8402
T8412
cf_saccharide
cf_fatty_acid
bacteriocin
nrps
t1pks-nrps
t1pks
transatpks-nrps
ectoine
transatpks
terpene-arylpolyene
siderophore
butyrolactone
arylpolyene
resorcinol
hserlactone
other
0
5
10
15
20
25
30
BGC counts

A. 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	SRA accession #
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		SRX7665675
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		SRX7665685
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		SRX7665686
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		SRX7665676
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		SRX7665684
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		SRX7665687
10	KP3700G	PMS-3700M	*Kuphus polythalamius* 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		SRX7665688
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		SRX7665689
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		SRX7665690
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	154360930	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	144540774	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	180149584	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	159707004	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	-	SRX7665691
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	-	SRX7665677

22	NR03	-	*Neoteredo reynei*	Coroa grande Mangrove - Sepetiba bay, Rio de Janeiro State, BR	22.9081670° S 43.8756390° W	CEGENBIO	Illumina MiSeq	SPAdes	13078802	309408486	414159	769	38.2	-	SRX7665678
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		SRX7665679
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		SRX7665680
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		SRX7665681
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		SRX7665682
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		SRX7665683

B: 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:	50.81	2513237099

												11/176.4 kbp		
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	Teredinibacter waterburyi	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* specimen PMS-2133X	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
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

genome_ID1	genome_ID2	AF	gANI	edge_score=AF*gANI	genome_length
1133Y	1675L	0.88837	0.97601	0.867058	4416264
1133Y	991H	0.917913	0.976538	0.896377	4416264
1133Y	NR01_83_0.fasta	0.625963	0.931249	0.582927	4416264
1133Y	NR02_1_1.fasta	0.828551	0.924503	0.765998	4416264
1133Y	NR03_uc.fasta	0.627183	0.936548	0.587387	4416264
1133Y	T0609	0.880447	0.974621	0.858102	4416264
1133Y	T7901	0.901882	0.930208	0.838938	4416264
1133Y	T7902	0.892839	0.976096	0.871497	4416264
1133Y	T8402	0.89629	0.931583	0.834969	4416264
1133Y	T8412	0.896951	0.972772	0.872529	4416264
1133Y	T8415	0.890276	0.932962	0.830594	4416264
1133Y	T8513	0.903741	0.975097	0.881235	4416264
1133Y	T8602	0.897988	0.930602	0.835669	4416264
1133Y	TBF05_2_0.fasta	0.818783	0.968715	0.793167	4416264
1675L	1133Y	0.872475	0.975545	0.851139	4536378
1675L	991H	0.890956	0.98154	0.874509	4536378
1675L	NR01_83_0.fasta	0.633585	0.931447	0.590151	4536378
1675L	NR02_1_1.fasta	0.827928	0.919958	0.761659	4536378
1675L	NR03_uc.fasta	0.635214	0.935692	0.594365	4536378
1675L	T0609	0.887587	0.979341	0.86925	4536378
1675L	T7901	0.868763	0.923789	0.802554	4536378
1675L	T7902	0.905002	0.986385	0.89268	4536378
1675L	T8402	0.86934	0.925605	0.804665	4536378
1675L	T8412	0.877281	0.979048	0.8589	4536378
1675L	T8415	0.894459	0.924962	0.827341	4536378
1675L	T8513	0.892232	0.982003	0.876175	4536378
1675L	T8602	0.855345	0.924807	0.791029	4536378
1675L	TBF05_2_0.fasta	0.783679	0.97568	0.76462	4536378
2141T	KP2132G_543	0.667543	0.904162	0.603567	4217910
2719K	DM2722G_579	0.871057	0.994387	0.866168	4518363
2719K	DM2858G_1458	0.855756	0.991247	0.848266	4518363
2719K	DM3770G_2725	0.783067	0.976768	0.764875	4518363

2753L	BT2849G_1909	0.626956	0.988297	0.619619	5329059
2753L	DM2858G_3735	0.520819	0.985097	0.513057	5329059
991H	1133Y	0.888098	0.975928	0.86672	4574307
991H	1675L	0.88192	0.982037	0.866078	4574307
991H	NR01_83_0.fasta	0.622196	0.930411	0.578898	4574307
991H	NR02_1_1.fasta	0.8166	0.917586	0.749301	4574307
991H	NR03_uc.fasta	0.632584	0.933684	0.590634	4574307
991H	T0609	0.86053	0.983114	0.845999	4574307
991H	T7901	0.873763	0.923005	0.806488	4574307
991H	T7902	0.894225	0.981671	0.877835	4574307
991H	T8402	0.880146	0.923846	0.813119	4574307
991H	T8412	0.89513	0.979792	0.877041	4574307
991H	T8415	0.869846	0.923357	0.803178	4574307
991H	T8513	0.885041	0.983463	0.870405	4574307
991H	T8602	0.85937	0.922279	0.792579	4574307
991H	TBF05_2_0.fasta	0.792542	0.976231	0.773704	4574307
BS08	BSG2_2_0.fasta	0.761704	0.992799	0.756219	4165124
BSC2.fasta	BSG1_1_1.fasta	0.87084	0.965521	0.840814	4577055
BSC2.fasta	BSG3_2_0.fasta	0.724032	0.958843	0.694233	4577055
BSC2.fasta	BSG4_1_0.fasta	0.711961	0.970806	0.691176	4577055
BSG1_1_1.fasta	BSC2.fasta	0.645775	0.973085	0.628394	6666104
BSG2_2_0.fasta	BS08	0.920644	0.995137	0.916167	3513639
BSG2_2_1.fasta	BSC2.fasta	0.898617	0.9748	0.875972	2557003
BSG2_2_4.fasta	BSC2.fasta	0.945788	0.976468	0.923532	323727
BSG2_2_9.fasta	BSC2.fasta	0.904797	0.980741	0.887372	137905
BSG3_2_0.fasta	BSC2.fasta	0.593708	0.973826	0.578168	6133323
BSG4_1_0.fasta	BSC2.fasta	0.861395	0.974795	0.839684	3869091
BT2771G_1251	2753L	0.737287	0.985666	0.726719	931086
BT2771G_1266	2753L	0.959437	0.99215	0.951905	1872123
BT2771G_2629	2753L	0.903905	0.991576	0.896291	2667810
BT2849G_1158	2753L	0.784775	0.9885	0.77575	1191456
BT2849G_1418	2753L	0.990501	0.992228	0.982803	16107
BT2849G_1523	2753L	0.540321	0.985773	0.532634	21465

BT2849G_1577	2753L	0.711287	0.963563	0.68537	36732
BT2849G_1909	2753L	0.971633	0.991404	0.963281	3503796
BT2849G_2869	2753L	0.830116	0.988927	0.820924	687471
BT3790G_1208	2753L	0.745048	0.982692	0.732153	402075
BT3790G_1493	2753L	0.768579	0.989095	0.760198	2051745
BT3790G_1981	2753L	0.552787	0.959462	0.530378	64305
BT3790G_2237	2753L	0.933243	0.991594	0.925398	3015375
BT3790G_3135	2753L	0.626175	0.981932	0.614861	732126
DM2722G_1447	1133Y	0.869977	0.935091	0.813508	2119464
DM2722G_1447	1675L	0.863317	0.928096	0.801241	2119464
DM2722G_1447	991H	0.865791	0.927151	0.802719	2119464
DM2722G_1447	T0609	0.859012	0.928516	0.797606	2119464
DM2722G_1447	T7901	0.891632	0.983404	0.876834	2119464
DM2722G_1447	T7902	0.872281	0.927808	0.809309	2119464
DM2722G_1447	T8402	0.891799	0.981492	0.875294	2119464
DM2722G_1447	T8412	0.863001	0.926899	0.799915	2119464
DM2722G_1447	T8415	0.885098	0.981571	0.868787	2119464
DM2722G_1447	T8513	0.864245	0.927157	0.801291	2119464
DM2722G_1447	T8602	0.895983	0.985196	0.882719	2119464
DM2722G_1691	2753L	0.860087	0.991096	0.852429	2544396
DM2722G_1870	1133Y	0.873943	0.927325	0.810429	2047428
DM2722G_1870	1675L	0.87697	0.921825	0.808413	2047428
DM2722G_1870	991H	0.877267	0.920066	0.807144	2047428
DM2722G_1870	T0609	0.878737	0.922672	0.810786	2047428
DM2722G_1870	T7901	0.914987	0.981546	0.898102	2047428
DM2722G_1870	T7902	0.871213	0.920104	0.801607	2047428
DM2722G_1870	T8402	0.900602	0.980756	0.883271	2047428
DM2722G_1870	T8412	0.867036	0.919955	0.797634	2047428
DM2722G_1870	T8415	0.895129	0.981002	0.878123	2047428
DM2722G_1870	T8513	0.861727	0.920596	0.793302	2047428
DM2722G_1870	T8602	0.905362	0.982463	0.889485	2047428
DM2722G_3144	2753L	0.682074	0.989466	0.674889	3191535
DM2722G_497	2719K	0.707764	0.988715	0.699777	515829

DM2722G_579	2719K	0.986046	0.99807	0.984143	4146420
DM2858G_1105	1133Y	0.620173	0.946415	0.586941	1252896
DM2858G_1105	1675L	0.633836	0.935645	0.593045	1252896
DM2858G_1105	991H	0.619454	0.931833	0.577228	1252896
DM2858G_1105	T0609	0.603874	0.934476	0.564306	1252896
DM2858G_1105	T7901	0.64117	0.976821	0.626308	1252896
DM2858G_1105	T7902	0.611744	0.933045	0.570785	1252896
DM2858G_1105	T8402	0.643978	0.973645	0.627006	1252896
DM2858G_1105	T8412	0.605823	0.931875	0.564551	1252896
DM2858G_1105	T8415	0.63788	0.974056	0.621331	1252896
DM2858G_1105	T8513	0.60504	0.93013	0.562766	1252896
DM2858G_1105	T8602	0.654028	0.976458	0.638631	1252896
DM2858G_1458	2719K	0.95337	0.996185	0.949733	4319439
DM2858G_2488	1133Y	0.527697	0.915996	0.483368	31086
DM2858G_2488	1675L	0.509168	0.922921	0.469922	31086
DM2858G_2488	991H	0.519494	0.912874	0.474233	31086
DM2858G_2488	T7901	0.565238	0.959991	0.542623	31086
DM2858G_2488	T7902	0.541884	0.911428	0.493888	31086
DM2858G_2488	T8402	0.539761	0.97622	0.526925	31086
DM2858G_2488	T8412	0.533681	0.914949	0.488291	31086
DM2858G_2488	T8415	0.570546	0.97412	0.55578	31086
DM2858G_2488	T8602	0.569774	0.971827	0.553722	31086
DM2858G_2501	2719K	0.701329	0.98137	0.688263	411372
DM2858G_2907	1133Y	0.630511	0.914008	0.576292	2150778
DM2858G_2907	1675L	0.623165	0.90994	0.567043	2150778
DM2858G_2907	991H	0.622968	0.909982	0.56689	2150778
DM2858G_2907	T0609	0.626758	0.910536	0.570686	2150778
DM2858G_2907	T7901	0.652576	0.972146	0.634399	2150778
DM2858G_2907	T7902	0.624774	0.908152	0.56739	2150778
DM2858G_2907	T8402	0.648725	0.972033	0.630582	2150778
DM2858G_2907	T8412	0.630313	0.9102	0.573711	2150778
DM2858G_2907	T8415	0.646015	0.972657	0.628351	2150778
DM2858G_2907	T8513	0.618733	0.909773	0.562907	2150778

DM2858G_2907	T8602	0.641596	0.972786	0.624136	2150778
DM2858G_3045	1133Y	0.919354	0.933049	0.857802	1967640
DM2858G_3045	1675L	0.914358	0.926672	0.84731	1967640
DM2858G_3045	991H	0.916906	0.926561	0.849569	1967640
DM2858G_3045	T0609	0.916133	0.927309	0.849538	1967640
DM2858G_3045	T7901	0.932526	0.984445	0.918021	1967640
DM2858G_3045	T7902	0.907632	0.926104	0.840562	1967640
DM2858G_3045	T8402	0.928931	0.982109	0.912311	1967640
DM2858G_3045	T8412	0.910497	0.925235	0.842424	1967640
DM2858G_3045	T8415	0.928217	0.981995	0.911504	1967640
DM2858G_3045	T8513	0.908572	0.925417	0.840808	1967640
DM2858G_3045	T8602	0.934802	0.986011	0.921725	1967640
DM2858G_3735	2753L	0.861203	0.991039	0.853486	3409581
DM3770G_1109	T8415	0.528894	0.9172	0.485102	13290
DM3770G_1432	T7901	0.515957	0.973758	0.502417	771279
DM3770G_2006	1133Y	0.936908	0.926142	0.86771	1755333
DM3770G_2006	1675L	0.936995	0.924306	0.86607	1755333
DM3770G_2006	991H	0.942717	0.923133	0.870253	1755333
DM3770G_2006	T0609	0.93762	0.923223	0.865632	1755333
DM3770G_2006	T7901	0.954063	0.985646	0.940368	1755333
DM3770G_2006	T7902	0.937723	0.922307	0.864868	1755333
DM3770G_2006	T8402	0.953564	0.983361	0.937698	1755333
DM3770G_2006	T8412	0.934559	0.92214	0.861794	1755333
DM3770G_2006	T8415	0.956646	0.983846	0.941192	1755333
DM3770G_2006	T8513	0.93782	0.922223	0.864879	1755333
DM3770G_2006	T8602	0.948791	0.987314	0.936755	1755333
DM3770G_2725	2719K	0.915257	0.985789	0.90225	4824342
DM3770G_2751	1133Y	0.892276	0.947536	0.845464	406707
DM3770G_2751	1675L	0.863191	0.938519	0.810121	406707
DM3770G_2751	991H	0.875938	0.938369	0.821953	406707
DM3770G_2751	T0609	0.876034	0.938738	0.822366	406707
DM3770G_2751	T7901	0.893228	0.974315	0.870285	406707
DM3770G_2751	T7902	0.849081	0.937778	0.796249	406707

DM3770G_2751	T8402	0.882879	0.971073	0.85734	406707
DM3770G_2751	T8412	0.869786	0.936096	0.814203	406707
DM3770G_2751	T8415	0.885077	0.968797	0.85746	406707
DM3770G_2751	T8513	0.85316	0.936312	0.798824	406707
DM3770G_2751	T8602	0.901799	0.973795	0.878167	406707
DM3770G_2901	2753L	0.791226	0.988497	0.782125	105042
DM3770G_2983	1133Y	0.741582	0.926188	0.686844	1267746
DM3770G_2983	1675L	0.710092	0.918912	0.652512	1267746
DM3770G_2983	991H	0.734198	0.918836	0.674608	1267746
DM3770G_2983	T0609	0.722201	0.918436	0.663295	1267746
DM3770G_2983	T7901	0.777272	0.981042	0.762536	1267746
DM3770G_2983	T7902	0.721815	0.918004	0.662629	1267746
DM3770G_2983	T8402	0.777461	0.980031	0.761936	1267746
DM3770G_2983	T8412	0.735069	0.919615	0.67598	1267746
DM3770G_2983	T8415	0.74838	0.98179	0.734752	1267746
DM3770G_2983	T8513	0.723786	0.918056	0.664476	1267746
DM3770G_2983	T8602	0.761282	0.980425	0.74638	1267746
DM3770G_3242	2753L	0.770013	0.980837	0.755257	37476
DM3770G_3460	T8602	0.521027	0.958985	0.499657	36453
DM3770G_5	2753L	0.51835	0.986173	0.511183	276537
DM3770G_615	1133Y	0.870698	0.943556	0.821552	1408404
DM3770G_615	1675L	0.879922	0.934937	0.822672	1408404
DM3770G_615	991H	0.878307	0.933908	0.820258	1408404
DM3770G_615	T0609	0.860142	0.935645	0.804788	1408404
DM3770G_615	T7901	0.90391	0.980828	0.88658	1408404
DM3770G_615	T7902	0.872341	0.932991	0.813886	1408404
DM3770G_615	T8402	0.904013	0.979725	0.885684	1408404
DM3770G_615	T8412	0.849266	0.932658	0.792075	1408404
DM3770G_615	T8415	0.908096	0.978877	0.888914	1408404
DM3770G_615	T8513	0.864238	0.933212	0.806517	1408404
DM3770G_615	T8602	0.915869	0.982445	0.899791	1408404
DM3770G_994	2719K	0.553608	0.964841	0.534144	677700
KP2132G_2024	2141T	0.951036	0.96525	0.917987	236499

KP2132G_487	2141T	0.751529	0.947831	0.712322	2396295
KP2132G_543	2141T	0.930654	0.947829	0.882101	7410314
KP2132G_930	2141T	0.913836	0.969131	0.885627	15633
KP2133G_110	2141T	0.970586	0.978075	0.949306	184775
KP2133G_12	2141T	0.852426	0.972543	0.829021	906223
KP2133G_1401	2141T	0.991723	0.918486	0.910884	36851
KP2133G_1537	2141T	0.937146	0.974138	0.91291	1388649
KP2133G_1802	2141T	0.885012	0.971645	0.859917	96236
KP2133G_407	2141T	0.514411	0.963889	0.495835	424581
KP2133G_561	2141T	0.847441	0.979638	0.830185	20805
KP2133G_574	2141T	0.956468	0.973746	0.931357	2478081
KP2133G_581	2141T	0.968286	0.976388	0.945423	52311
KP2133G_742	2141T	0.846229	0.972483	0.822943	860714
KP3700G_1264	2141T	0.916918	0.979339	0.897974	5185779
KP3700G_1558	2141T	0.715557	0.973669	0.696716	1902657
KP3700G_2285	2141T	0.523456	0.953913	0.499331	94260
KP3700G_254	2141T	0.65172	0.96095	0.62627	67506
NR01_83_0.fasta	1133Y	0.67502	0.946155	0.638674	6971531
NR01_83_0.fasta	1675L	0.6816	0.945046	0.644143	6971531
NR01_83_0.fasta	991H	0.694159	0.946682	0.657148	6971531
NR01_83_0.fasta	T0609	0.660568	0.946208	0.625035	6971531
NR01_83_0.fasta	T7901	0.689513	0.945465	0.65191	6971531
NR01_83_0.fasta	T7902	0.686353	0.946113	0.649367	6971531
NR01_83_0.fasta	T8402	0.687244	0.944579	0.649156	6971531
NR01_83_0.fasta	T8412	0.693179	0.945101	0.655124	6971531
NR01_83_0.fasta	T8415	0.679432	0.944509	0.64173	6971531
NR01_83_0.fasta	T8513	0.713256	0.953623	0.680177	6971531
NR01_83_0.fasta	T8602	0.674679	0.944171	0.637012	6971531
NR01_uc.fasta	1133Y	0.89634	0.937844	0.840627	1878871
NR01_uc.fasta	1675L	0.900039	0.932247	0.839059	1878871
NR01_uc.fasta	991H	0.896425	0.932008	0.835475	1878871
NR01_uc.fasta	T0609	0.894596	0.932416	0.834136	1878871
NR01_uc.fasta	T7901	0.908853	0.979432	0.89016	1878871

NR01_uc.fasta	T7902	0.898114	0.931294	0.836408	1878871
NR01_uc.fasta	T8402	0.912015	0.97896	0.892826	1878871
NR01_uc.fasta	T8412	0.891515	0.931243	0.830217	1878871
NR01_uc.fasta	T8415	0.912593	0.97762	0.892169	1878871
NR01_uc.fasta	T8513	0.894387	0.932666	0.834164	1878871
NR01_uc.fasta	T8602	0.915191	0.978577	0.895585	1878871
NR02_1_1.fasta	1133Y	0.668416	0.931331	0.622517	6001350
NR02_1_1.fasta	1675L	0.679296	0.925157	0.628455	6001350
NR02_1_1.fasta	991H	0.683963	0.924842	0.632558	6001350
NR02_1_1.fasta	T0609	0.658962	0.925538	0.609894	6001350
NR02_1_1.fasta	T7901	0.692028	0.983431	0.680562	6001350
NR02_1_1.fasta	T7902	0.674652	0.92368	0.623163	6001350
NR02_1_1.fasta	T8402	0.69478	0.979676	0.680659	6001350
NR02_1_1.fasta	T8412	0.675651	0.923138	0.623719	6001350
NR02_1_1.fasta	T8415	0.69717	0.980238	0.683393	6001350
NR02_1_1.fasta	T8513	0.682099	0.924662	0.630711	6001350
NR02_1_1.fasta	T8602	0.67917	0.981851	0.666844	6001350
NR03_1_5.fasta	1133Y	0.876318	0.936051	0.820278	62313
NR03_1_5.fasta	1675L	0.910404	0.960303	0.874264	62313
NR03_1_5.fasta	991H	0.932117	0.959799	0.894645	62313
NR03_1_5.fasta	T0609	0.818304	0.954325	0.780928	62313
NR03_1_5.fasta	T7901	0.955996	0.947273	0.905589	62313
NR03_1_5.fasta	T7902	0.897549	0.955068	0.85722	62313
NR03_1_5.fasta	T8402	0.880506	0.973973	0.857589	62313
NR03_1_5.fasta	T8412	0.876318	0.940355	0.82405	62313
NR03_1_5.fasta	T8415	0.880506	0.973992	0.857606	62313
NR03_1_5.fasta	T8513	0.824804	0.947214	0.781266	62313
NR03_1_5.fasta	T8602	0.862067	0.958841	0.826585	62313
NR03_3_0.fasta	1133Y	0.668093	0.935554	0.625037	1713843
NR03_3_0.fasta	1675L	0.681904	0.932979	0.636202	1713843
NR03_3_0.fasta	991H	0.673016	0.932529	0.627607	1713843
NR03_3_0.fasta	T0609	0.665706	0.93191	0.620378	1713843
NR03_3_0.fasta	T7901	0.675478	0.965013	0.651845	1713843

NR03_3_0.fasta	T7902	0.673639	0.932215	0.627976	1713843
NR03_3_0.fasta	T8402	0.666042	0.962142	0.640827	1713843
NR03_3_0.fasta	T8412	0.666094	0.931251	0.620301	1713843
NR03_3_0.fasta	T8415	0.669292	0.963475	0.644846	1713843
NR03_3_0.fasta	T8513	0.684772	0.933819	0.639453	1713843
NR03_3_0.fasta	T8602	0.661184	0.963555	0.637087	1713843
NR03_uc.fasta	1133Y	0.613876	0.946813	0.581226	7380431
NR03_uc.fasta	1675L	0.620773	0.945993	0.587247	7380431
NR03_uc.fasta	991H	0.638711	0.945861	0.604132	7380431
NR03_uc.fasta	T0609	0.607121	0.945044	0.573756	7380431
NR03_uc.fasta	T7901	0.622761	0.95804	0.59663	7380431
NR03_uc.fasta	T7902	0.628333	0.945183	0.59389	7380431
NR03_uc.fasta	T8402	0.638135	0.957549	0.611046	7380431
NR03_uc.fasta	T8412	0.636491	0.943442	0.600492	7380431
NR03_uc.fasta	T8415	0.628302	0.956369	0.600889	7380431
NR03_uc.fasta	T8513	0.642384	0.946389	0.607945	7380431
NR03_uc.fasta	T8602	0.615265	0.956948	0.588777	7380431
T0609	1133Y	0.885222	0.974391	0.862552	4392680
T0609	1675L	0.90846	0.978659	0.889073	4392680
T0609	991H	0.896712	0.982882	0.881362	4392680
T0609	NR01_83_0.fasta	0.627597	0.93272	0.585372	4392680
T0609	NR02_1_1.fasta	0.827633	0.920637	0.76195	4392680
T0609	NR03_uc.fasta	0.636922	0.935509	0.595846	4392680
T0609	T7901	0.873915	0.924175	0.80765	4392680
T0609	T7902	0.894391	0.981676	0.878002	4392680
T0609	T8402	0.876731	0.926249	0.812071	4392680
T0609	T8412	0.895122	0.97822	0.875626	4392680
T0609	T8415	0.884963	0.923698	0.817439	4392680
T0609	T8513	0.892209	0.982359	0.87647	4392680
T0609	T8602	0.870858	0.923299	0.804062	4392680
T0609	TBF05_2_0.fasta	0.809401	0.97434	0.788632	4392680
T7901	1133Y	0.889122	0.930659	0.827469	4485681
T7901	1675L	0.874732	0.924164	0.808396	4485681

T7901	991H	0.891699	0.923504	0.823488	4485681
T7901	NR01_83_0.fasta	0.631319	0.946123	0.597305	4485681
T7901	NR02_1_1.fasta	0.840667	0.979563	0.823486	4485681
T7901	NR03_uc.fasta	0.626208	0.95685	0.599187	4485681
T7901	T0609	0.857654	0.924544	0.792939	4485681
T7901	T7902	0.8819	0.922027	0.813136	4485681
T7901	T8402	0.910335	0.981911	0.893868	4485681
T7901	T8412	0.880605	0.921929	0.811855	4485681
T7901	T8415	0.902292	0.980763	0.884935	4485681
T7901	T8513	0.87447	0.92361	0.807669	4485681
T7901	T8602	0.901004	0.982316	0.885071	4485681
T7901	TBF05_2_0.fasta	0.796251	0.918111	0.731047	4485681
T7902	1133Y	0.846907	0.976243	0.826787	4679711
T7902	1675L	0.902333	0.985913	0.889622	4679711
T7902	991H	0.879235	0.982723	0.864044	4679711
T7902	NR01_83_0.fasta	0.616936	0.927803	0.572395	4679711
T7902	NR02_1_1.fasta	0.804973	0.917494	0.738558	4679711
T7902	NR03_uc.fasta	0.625576	0.933852	0.584195	4679711
T7902	T0609	0.853667	0.979927	0.836531	4679711
T7902	T7901	0.853014	0.921915	0.786406	4679711
T7902	T8402	0.881667	0.921382	0.812352	4679711
T7902	T8412	0.887045	0.979357	0.868734	4679711
T7902	T8415	0.875878	0.921236	0.80689	4679711
T7902	T8513	0.882228	0.981673	0.866059	4679711
T7902	T8602	0.839603	0.922374	0.774428	4679711
T7902	TBF05_2_0.fasta	0.770507	0.976281	0.752231	4679711
T8402	1133Y	0.880803	0.931718	0.82066	4501956
T8402	1675L	0.871153	0.925977	0.806668	4501956
T8402	991H	0.889551	0.925788	0.823536	4501956
T8402	NR01_83_0.fasta	0.624453	0.944341	0.589697	4501956
T8402	NR02_1_1.fasta	0.838445	0.976034	0.818351	4501956
T8402	NR03_uc.fasta	0.640692	0.955863	0.612414	4501956
T8402	T0609	0.855565	0.926143	0.792376	4501956

T8402	T7901	0.911137	0.981586	0.894359	4501956
T8402	T7902	0.891047	0.9231	0.822525	4501956
T8402	T8412	0.891846	0.923396	0.823527	4501956
T8402	T8415	0.912321	0.979713	0.893813	4501956
T8402	T8513	0.879714	0.925564	0.814232	4501956
T8402	T8602	0.898003	0.980678	0.880652	4501956
T8402	TBF05_2_0.fasta	0.788936	0.918738	0.724825	4501956
T8412	1133Y	0.890765	0.972143	0.865951	4467551
T8412	1675L	0.889963	0.979045	0.871314	4467551
T8412	991H	0.920565	0.98011	0.902255	4467551
T8412	NR01_83_0.fasta	0.635071	0.930336	0.590829	4467551
T8412	NR02_1_1.fasta	0.824026	0.916729	0.755409	4467551
T8412	NR03_uc.fasta	0.638637	0.933147	0.595942	4467551
T8412	T0609	0.882977	0.978528	0.864018	4467551
T8412	T7901	0.891578	0.921872	0.821921	4467551
T8412	T7902	0.917881	0.979589	0.899146	4467551
T8412	T8402	0.901942	0.923388	0.832842	4467551
T8412	T8415	0.883512	0.921931	0.814537	4467551
T8412	T8513	0.918563	0.979996	0.900188	4467551
T8412	T8602	0.8739	0.922244	0.805949	4467551
T8412	TBF05_2_0.fasta	0.816165	0.974634	0.795462	4467551
T8415	1133Y	0.87938	0.93364	0.821024	4476275
T8415	1675L	0.898434	0.926844	0.832708	4476275
T8415	991H	0.884153	0.925735	0.818491	4476275
T8415	NR01_83_0.fasta	0.636397	0.946789	0.602534	4476275
T8415	NR02_1_1.fasta	0.852261	0.977522	0.833104	4476275
T8415	NR03_uc.fasta	0.641977	0.956194	0.613855	4476275
T8415	T0609	0.86685	0.924166	0.801113	4476275
T8415	T7901	0.906795	0.981376	0.889907	4476275
T8415	T7902	0.884674	0.924949	0.818278	4476275
T8415	T8402	0.91731	0.981868	0.900677	4476275
T8415	T8412	0.877572	0.924005	0.810881	4476275
T8415	T8513	0.876244	0.925177	0.810681	4476275

T8415	T8602	0.889734	0.982798	0.874429	4476275
T8415	TBF05_2_0.fasta	0.779922	0.91913	0.71685	4476275
T8513	1133Y	0.882059	0.974339	0.859424	4559547
T8513	1675L	0.889501	0.981425	0.872979	4559547
T8513	991H	0.894236	0.983704	0.879664	4559547
T8513	NR01_83_0.fasta	0.649385	0.938047	0.609154	4559547
T8513	NR02_1_1.fasta	0.817981	0.917444	0.750452	4559547
T8513	NR03_uc.fasta	0.635693	0.936753	0.595487	4559547
T8513	T0609	0.861942	0.982137	0.846545	4559547
T8513	T7901	0.870331	0.92317	0.803463	4559547
T8513	T7902	0.895824	0.981855	0.879569	4559547
T8513	T8402	0.873841	0.925228	0.808502	4559547
T8513	T8412	0.901188	0.980463	0.883581	4559547
T8513	T8415	0.863506	0.924296	0.798135	4559547
T8513	T8602	0.862103	0.920981	0.79398	4559547
T8513	TBF05_2_0.fasta	0.798762	0.975963	0.779562	4559547
T8602	1133Y	0.892922	0.931257	0.83154	4444880
T8602	1675L	0.866881	0.924539	0.801465	4444880
T8602	991H	0.885145	0.922655	0.816683	4444880
T8602	NR01_83_0.fasta	0.62017	0.945379	0.586296	4444880
T8602	NR02_1_1.fasta	0.835341	0.977696	0.81671	4444880
T8602	NR03_uc.fasta	0.626023	0.955289	0.598033	4444880
T8602	T0609	0.862304	0.923203	0.796082	4444880
T8602	T7901	0.909376	0.982195	0.893185	4444880
T8602	T7902	0.87942	0.922876	0.811596	4444880
T8602	T8402	0.910518	0.9807	0.892945	4444880
T8602	T8412	0.877947	0.922503	0.809909	4444880
T8602	T8415	0.896389	0.981936	0.880197	4444880
T8602	T8513	0.881462	0.921479	0.812249	4444880
T8602	TBF05_2_0.fasta	0.80329	0.917094	0.736692	4444880
TBF02_1_1.fasta	1162T	0.960455	0.984263	0.94534	450549
TBF02_3_0.fasta	1133Y	0.8044	0.930773	0.748714	3234669
TBF02_3_0.fasta	1675L	0.788952	0.924999	0.72978	3234669

TBF02_3_0.fasta	991H	0.808884	0.924883	0.748123	3234669
TBF02_3_0.fasta	T0609	0.783988	0.925451	0.725542	3234669
TBF02_3_0.fasta	T7901	0.820008	0.976303	0.800576	3234669
TBF02_3_0.fasta	T7902	0.794959	0.924469	0.734915	3234669
TBF02_3_0.fasta	T8402	0.813269	0.975412	0.793272	3234669
TBF02_3_0.fasta	T8412	0.802364	0.923339	0.740854	3234669
TBF02_3_0.fasta	T8415	0.798753	0.975965	0.779555	3234669
TBF02_3_0.fasta	T8513	0.798417	0.922979	0.736922	3234669
TBF02_3_0.fasta	T8602	0.813756	0.978318	0.796112	3234669
TBF03_2_0.fasta	1133Y	0.77877	0.972509	0.757361	4012364
TBF03_2_0.fasta	1675L	0.769773	0.97976	0.754193	4012364
TBF03_2_0.fasta	991H	0.792989	0.980324	0.777386	4012364
TBF03_2_0.fasta	T0609	0.75793	0.979196	0.742162	4012364
TBF03_2_0.fasta	T7901	0.77334	0.919818	0.711332	4012364
TBF03_2_0.fasta	T7902	0.784174	0.980219	0.768662	4012364
TBF03_2_0.fasta	T8402	0.768395	0.921958	0.708428	4012364
TBF03_2_0.fasta	T8412	0.785248	0.97951	0.769158	4012364
TBF03_2_0.fasta	T8415	0.752277	0.921598	0.693297	4012364
TBF03_2_0.fasta	T8513	0.781884	0.981158	0.767152	4012364
TBF03_2_0.fasta	T8602	0.75842	0.919919	0.697685	4012364
TBF05_2_0.fasta	1133Y	0.92992	0.973602	0.905372	4056823
TBF05_2_0.fasta	1675L	0.91262	0.980719	0.895024	4056823
TBF05_2_0.fasta	991H	0.935601	0.981484	0.918277	4056823
TBF05_2_0.fasta	T0609	0.915587	0.980034	0.897306	4056823
TBF05_2_0.fasta	T7901	0.920769	0.92287	0.84975	4056823
TBF05_2_0.fasta	T7902	0.923265	0.981124	0.905837	4056823
TBF05_2_0.fasta	T8402	0.912109	0.923725	0.842538	4056823
TBF05_2_0.fasta	T8412	0.933926	0.98047	0.915686	4056823
TBF05_2_0.fasta	T8415	0.90229	0.923779	0.833517	4056823
TBF05_2_0.fasta	T8513	0.932749	0.982141	0.916091	4056823
TBF05_2_0.fasta	T8602	0.918734	0.921887	0.846969	4056823
TBF05_uc.fasta	1133Y	0.659554	0.965507	0.636804	198111
TBF05_uc.fasta	1675L	0.646047	0.977428	0.631464	198111

TBF05_uc.fasta	991H	0.640762	0.978234	0.626815	198111
TBF05_uc.fasta	T0609	0.667929	0.972741	0.649722	198111
TBF05_uc.fasta	T7901	0.634311	0.91599	0.581023	198111
TBF05_uc.fasta	T7902	0.630964	0.977544	0.616795	198111
TBF05_uc.fasta	T8402	0.668655	0.919369	0.614741	198111
TBF05_uc.fasta	T8412	0.665173	0.966466	0.642867	198111
TBF05_uc.fasta	T8415	0.635492	0.916043	0.582138	198111
TBF05_uc.fasta	T8513	0.631192	0.97748	0.616978	198111
TBF05_uc.fasta	T8602	0.644336	0.912996	0.588276	198111
TBF07_1_1.fasta	1162T	0.895614	0.984754	0.881959	743952
TBF09_17_0.fasta	1133Y	0.953758	0.971264	0.926351	1861029
TBF09_17_0.fasta	1675L	0.939717	0.978883	0.919873	1861029
TBF09_17_0.fasta	991H	0.966973	0.979451	0.947103	1861029
TBF09_17_0.fasta	T0609	0.930403	0.97855	0.910446	1861029
TBF09_17_0.fasta	T7901	0.943027	0.922276	0.869731	1861029
TBF09_17_0.fasta	T7902	0.956445	0.979964	0.937282	1861029
TBF09_17_0.fasta	T8402	0.937665	0.924468	0.866841	1861029
TBF09_17_0.fasta	T8412	0.96116	0.979005	0.94098	1861029
TBF09_17_0.fasta	T8415	0.920284	0.923824	0.85018	1861029
TBF09_17_0.fasta	T8513	0.955093	0.979827	0.935826	1861029
TBF09_17_0.fasta	T8602	0.934865	0.921081	0.861086	1861029
TBF09_2_0.fasta	1162T	0.943394	0.981959	0.926374	294653

Query sequence	Subject sequence	%Ident	length	evalue	qcovs	qlen	slen
NODE_14347_length_5660_cov_6.146055	NODE_31934_length_3870_cov_17.413977	94.28	3883	0	68	5660	3870
NODE_23307_length_4536_cov_6.299660	NODE_23527_length_4515_cov_3.470414	94.37	3377	0	83	4536	4515
NODE_23527_length_4515_cov_3.470414	NODE_23307_length_4536_cov_6.299660	94.37	3377	0	83	4515	4536
NODE_25141_length_4367_cov_9.751766	NODE_53552_length_2897_cov_8.209654	95.48	2897	0	66	4367	2897
NODE_31934_length_3870_cov_17.413977	NODE_14347_length_5660_cov_6.146055	94.28	3883	0	100	3870	5660
NODE_35626_length_3655_cov_13.375212	NODE_36097_length_3631_cov_8.919088	94.18	3661	0	100	3655	3631
NODE_36097_length_3631_cov_8.919088	NODE_35626_length_3655_cov_13.375212	94.18	3661	0	100	3631	3655
NODE_38563_length_3503_cov_8.494973	NODE_39659_length_3448_cov_13.046889	92.9	3465	0	98	3503	3448
NODE_39659_length_3448_cov_13.046889	NODE_38563_length_3503_cov_8.494973	92.9	3464	0	100	3448	3503
NODE_40202_length_3424_cov_11.970027	NODE_67412_length_2505_cov_5.815856	96.77	2507	0	73	3424	2505
NODE_41049_length_3384_cov_11.921851	NODE_66289_length_2532_cov_14.121526	97.67	2532	0	75	3384	2532
NODE_51053_length_2982_cov_12.870675	NODE_73070_length_2374_cov_7.126054	96.46	2374	0	80	2982	2374
NODE_51562_length_2964_cov_4.371790	NODE_62798_length_2623_cov_3.187450	95.89	2629	0	88	2964	2623
NODE_53552_length_2897_cov_8.209654	NODE_25141_length_4367_cov_9.751766	95.48	2897	0	100	2897	4367
NODE_56563_length_2800_cov_13.415454	NODE_67976_length_2491_cov_5.948101	95.54	2488	0	89	2800	2491
NODE_60745_length_2677_cov_6.553208	NODE_61584_length_2655_cov_4.045383	96.95	1806	0	67	2677	2655
NODE_61584_length_2655_cov_4.045383	NODE_60745_length_2677_cov_6.553208	96.95	1806	0	68	2655	2677
NODE_62798_length_2623_cov_3.187450	NODE_51562_length_2964_cov_4.371790	95.89	2629	0	99	2623	2964
NODE_62828_length_2621_cov_13.203200	NODE_68217_length_2485_cov_15.815990	94	2485	0	95	2621	2485
NODE_64237_length_2585_cov_4.464692	NODE_79312_length_2246_cov_3.968471	93.8	1840	0	71	2585	2246
NODE_66289_length_2532_cov_14.121526	NODE_41049_length_3384_cov_11.921851	97.67	2532	0	100	2532	3384
NODE_66682_length_2523_cov_4.220649	NODE_70658_length_2428_cov_4.239272	95.65	2025	0	80	2523	2428
NODE_67412_length_2505_cov_5.815856	NODE_40202_length_3424_cov_11.970027	96.77	2507	0	100	2505	3424
NODE_67878_length_2494_cov_4.392752	NODE_69000_length_2466_cov_4.622601	96.47	1614	0	65	2494	2466
NODE_67976_length_2491_cov_5.948101	NODE_56563_length_2800_cov_13.415454	95.54	2488	0	99	2491	2800
NODE_68217_length_2485_cov_15.815990	NODE_62828_length_2621_cov_13.203200	94	2485	0	100	2485	2621
NODE_69000_length_2466_cov_4.622601	NODE_67878_length_2494_cov_4.392752	96.47	1614	0	65	2466	2494
NODE_69500_length_2454_cov_10.399914	NODE_69501_length_2454_cov_9.504929	97.6	2454	0	100	2454	2454
NODE_69501_length_2454_cov_9.504929	NODE_69500_length_2454_cov_10.399914	97.6	2454	0	100	2454	2454
NODE_70658_length_2428_cov_4.239272	NODE_66682_length_2523_cov_4.220649	95.65	2025	0	83	2428	2523
NODE_72738_length_2381_cov_15.003540	NODE_73067_length_2374_cov_13.619174	97.31	2382	0	100	2381	2374
NODE_73028_length_2376_cov_2.580488	NODE_74717_length_2339_cov_8.961226	94.38	2009	0	84	2376	2339

NODE_73067_length_2374_cov_13.619174	NODE_72738_length_2381_cov_15.003540	97.31	2382	0	100	2374	2381
NODE_73070_length_2374_cov_7.126054	NODE_51053_length_2982_cov_12.870675	96.46	2374	0	100	2374	2982
NODE_74717_length_2339_cov_8.961226	NODE_73028_length_2376_cov_2.580488	94.38	2009	0	85	2339	2376
NODE_79312_length_2246_cov_3.968471	NODE_64237_length_2585_cov_4.464692	93.8	1840	0	81	2246	2585
NODE_81302_length_2208_cov_11.453282	NODE_86577_length_2114_cov_8.161565	95.36	2114	0	96	2208	2114
NODE_81429_length_2206_cov_8.507434	NODE_83204_length_2173_cov_5.881579	93.25	2207	0	100	2206	2173
NODE_83204_length_2173_cov_5.881579	NODE_81429_length_2206_cov_8.507434	93.25	2207	0	100	2173	2206
NODE_86577_length_2114_cov_8.161565	NODE_81302_length_2208_cov_11.453282	95.36	2114	0	100	2114	2208

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)