Discovery of an Antarctic ascidian-associated uncultivated *Verrucomicrobia* with antimelanoma palmerolide biosynthetic potential.

Alison E. Murray[a*], Chien-Chi Lo[b], Hajnalka E. Daligault[b], Nicole E. Avalon[c], Robert W. Read[a], Karen W. Davenport[b], Mary L. Higham[a], Yuliya Kunde[b], Armand E.K. Dichosa[b], Bill J. Baker[c*], Patrick S.G. Chain[b*]

[a] Division of Earth and Ecosystem Science, Desert Research Institute, Reno, Nevada, USA.

[b] Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico, USA.

[c] Department of Chemistry, University of South Florida, Tampa, Florida, USA.

*Corresponding authors:

Alison E. Murray, Division of Earth and Ecosystem Sciences, Desert Research Institute, Reno, NV 89512, USA. +1 775-673-7361. alison.murray@dri.edu.

Bill J. Baker, Department of Chemistry, University of South Florida, Tampa, FL 33620, USA. +1-813-974-1967, bjbaker@usf.edu

Patrick S.G. Chain, Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. +1-505-665-4019. pchain@lanl.gov

Abstract

The Antarctic marine ecosystem harbors a wealth of biological and chemical innovation that has risen in concert over millennia since the isolation of the continent and formation of the Antarctic circumpolar current. Scientific inquiry into the novelty of marine natural products produced by Antarctic benthic invertebrates led to the discovery of a bioactive macrolide, palmerolide A, that has specific activity against melanoma and holds considerable promise as an anticancer therapeutic. While this compound was isolated from the Antarctic ascidian *Synoicum adareanum*, its biosynthesis has since been hypothesized to be microbially mediated, given structural similarities to microbially-produced hybrid non-ribosomal peptide-polyketide macrolides. Here, we describe a metagenome-enabled investigation aimed at identifying the biosynthetic gene cluster (BGC) and palmerolide A-producing organism. A 74 Kbp candidate BGC encoding the multi-modular enzymatic machinery (hybrid Type I-*trans*-AT polyketide synthase-non-ribosomal peptide synthetase and tailoring functional domains) was identified and found to harbor key features predicted as necessary for palmerolide A biosynthesis. Surveys of ascidian microbiome samples targeting the candidate BGC revealed a high correlation between palmerolide-gene targets and a single 16S rRNA gene variant (R=0.83 – 0.99). Through repeated rounds of metagenome sequencing followed by binning contigs into metagenome-assembled genomes, we were able to retrieve a near-complete genome (10 contigs) of the BGC-producing organism, a novel verrucomicrobium within the *Opitutaceae* family that we propose here as *Candidatus* Synoicihabitans palmerolidicus. The refined genome assembly harbors five highly similar BGC copies, along with structural and functional features that shed light on the host-associated nature of this unique bacterium.

48 **Importance**

49 Palmerolide A has potential as a chemotherapeutic agent to target melanoma. We interrogated
50 the microbiome of the Antarctic ascidian, *Synoicum adareanum,* using a cultivation-independent
51 high-throughput sequencing and bioinformatic strategy. The metagenome-encoded biosynthetic
52 machinery predicted to produce palmerolide A was found to be associated with the genome of a
53 member of the *S. adareanum* core microbiome. Phylogenomic analysis suggests the organism
54 represents a new deeply-branching genus, *Candidatus* Synoicihabitans palmerolidicus, in the
55 *Opitutaceae* family of the Verrucomicrobia phylum. The *Ca*. S. palmerolidicus 4.29 Mb genome
56 encodes a repertoire of carbohydrate-utilizing and transport pathways enabling its ascidian-
57 associated lifestyle. The palmerolide-producer's genome also contains five distinct copies of the
58 large palmerolide biosynthetic gene cluster that may provide structural complexity of
59 palmerolide variants.

60

61 **Introduction**

62 Across the world's oceans, marine benthic invertebrates harbor a rich source of natural products
63 that serve metabolic and ecological roles in situ. These compounds provide a multitude of
64 medicinal and biotechnological applications to science, health and industry. The organisms
65 responsible for their biosynthesis are often unknown (1, 2). Increasingly, these metabolites,
66 especially in the polyketide class (*trans*-AT in particular), are found to be produced by microbial
67 counterparts associated with the invertebrate host (3-5). Invertebrates including sponges, corals,
68 and ascidians for example, harbor a wealth of diverse microbes, few of which have been
69 cultivated (e.g., 6-8). Genomic tools, in particular, are revealing biochemical pathways
70 potentially critical in the host-microbe associations (9). Microbes that form persistent mutualistic

71 (symbiotic) associations provide key roles in host ecology, such as provision of metabolic
72 requirements, production of adaptive features such as photoprotective pigments,
73 bioluminescence, or antifoulants, and biosynthesis of chemical defense agents.
74 Antarctic marine ecosystems harbor species-rich macrobenthic communities (10-12), which
75 have been the subject of natural products investigations over the past 30 years resulting in the
76 identification of > 600 metabolites (13). Initially, it was not known whether the same selective
77 pressures (namely predation and competition, e.g., 14) that operate in mid and low latitudes
78 would drive benthic organisms at the poles to create novel chemistry (15). However, this does
79 appear to be the case, and novel natural products have been discovered across algae, sponges,
80 corals, nudibranchs, echinoderms, bryozoans, ascidians, and increasingly amongst
81 microorganisms (16) for which the ecological roles have been deduced in a number of cases (13,
82 17). Studies of Antarctic benthic invertebrate-microbe associations however, pale in comparison
83 to studies at lower latitudes, yet the few studies that have been reported suggest these
84 associations (i) harbor an untapped reservoir of biological diversity (18-21) including fungi (22),
85 (ii) are host species-specific (23, 24), (iii) provide the host with sources of nitrogen and fixed
86 carbon (25) and (iv) have biosynthetic functional potential (26, 27).
87 This study was specifically motivated by our desire to understand the biosynthetic origins of
88 a natural product, palmerolide A, given its potent anticancer activity (28), that is found to be
89 associated with the polyclinid Antarctic ascidian, *Synoicum adareanum* (Fig. 1A and B).
90 Ascidians are known to be rich sources of bioactive natural products (9). They have been found
91 to harbor polyketide, terpenoid, peptide, alkaloid and a few other classes of natural products, of
92 which the majority have cytotoxic and/or antimicrobial activities. In addition to palmerolide A, a
93 few other natural products derived from Antarctic ascidians have been reported (29-31).

94 Ascidian-associated microbes responsible for natural product biosynthesis have been shown to
95 be affiliated with bacterial phyla including Actinobacteria (which dominates the recognized
96 diversity), Cyanobacteria, Firmicutes, Proteobacteria (both Alphaproteobacteria and
97 Gammaproteobacteria) and Verrucomicrobia in addition to many fungi (32, 33). Metagenome-
98 enabled studies have been key in linking natural products to the organisms producing them in a
99 number of cases, e.g. patellamide A and C to Cyanobacteria-affiliated *Prochloron* spp. (4), the
100 tetrahydroisoquinoline alkaloid ET-743 to Gammaproteobacteria-affiliated *Candidatus*
101 Endoecteinascidia frumentensis (34), patellazoles to Alphaproteobacteria-affiliated *Candidatus*
102 Endolissoclinum faulkneri (35), and mandelalides to Verrucomicrobia-affiliated *Candidatus*
103 Didemnitutus mandela (36). However, this is most certainly an under-representation of the
104 diversity of ascidian-associated microorganisms with capabilities for synthesizing bioactive
105 compounds, given the breadth of ascidian biodiversity (37). These linkages have been yet to be
106 investigated for Antarctic ascidians.
107 Palmerolide A has anticancer properties with selective activity against melanoma when
108 tested in the National Cancer Institute 60 cell-line panel (28). This result is of particular interest,
109 as there are few natural product therapeutics for this devastating form of cancer. Palmerolide A
110 inhibits vacuolar ATPases, which are highly expressed in metastatic melanoma. Given the
111 current level of understanding that macrolides often have microbial biosynthetic origins, that the
112 holobiont metagenome has biosynthetic potential (26) and a diverse, yet persistent core
113 microbiome is found in palmerolide A-containing *S. adareanum* (27) we have hypothesized that
114 a microbe associated with *S. adareanum* is responsible for the biosynthesis of palmerolide A.
115 The core microbiome of the palmerolide A-producing ascidian *S. adareanum* in samples
116 collected across the Anvers Island archipelago (n=63 samples; 27) is comprised of five bacterial

117 phyla including Proteobacteria (dominating the microbiome), Bacteroidetes, Nitrospirae,
118 Actinobacteria and Verrucomicrobia. A few candidate taxa in particular, were suggested to be
119 likely palmerolide A producers based on relative abundance and biosynthetic potential
120 determined by analysis of lineage-targeted biosynthetic capability (genera: *Microbulbifer*,
121 *Pseudovibrio*, *Hoeflea*, and the family *Opitutaceae* (27). This motivated interrogation of the *S.*
122 *adareanum* microbiome metagenome, with the goals of determining the metagenome-encoded
123 biosynthetic potential, identifying candidate palmerolide A biosynthetic gene cluster (BGC)(s)
124 and establishing the identity of the palmerolide A producing organism.
125

126 **Results and Discussion**

127 **Identification of a candidate palmerolide A biosynthetic gene cluster.** Microbial-enriched
128 fractions of *S. adareanum* metagenomic DNA sequence from 454 and Ion Proton next generation
129 sequencing (NGS) libraries (almost 18 billion bases in all) were assembled independently, then
130 merged, resulting in ~ 145 Mbp of assembled bases distributed over 86,387 contigs (referred to
131 as CoAssembly 1; supplemental material, Table S1). As the metagenome sequencing effort was
132 focused on identifying potential BGCs encoding the machinery to synthesize palmerolide A, the
133 initial steps of analysis specifically targeted those contigs in the assembly that were > 40 Kbp
134 (102 contigs representing 0.12 % of all contigs, which represent 7.07 % of all reads mapped), as
135 the size of the macrolide ring with 24 carbons would require a large number of polyketide
136 modules to be encoded. This large fragment subset of CoAssembly 1 was submitted to
137 antiSMASH v.3 (38) and more recently to v.5 (39). The results indicated a heterogeneous suite
138 of BGCs, including a bacteriocin, two non-ribosomal peptide synthetases (NRPS), two hybrid

139 NRPS -Type I PKS, two terpenes, and three *trans*-AT-PKS hybrid NRPS clusters (supplemental
140 material, Table S2).

141 We predicted several functional characteristics of the BGC that would be required for
142 palmerolide A biosynthesis which aided our analysis (see (40) for detailed retrobiosynthetic
143 analysis). This included evidence of a hybrid nonribosomal peptide-polyketide pathway and
144 enzymatic domains leading to placement of two distinct structural features of the polyketide
145 backbone, a carbamoyl transferase that appends a carbamate group at C-11 of the macrolide ring,
146 and an HMGCoA synthetase that inserts a methyl group on an acetate C-1 position of the
147 macrolide structure (C-25). The antiSMASH results indicated that two of the three predicted
148 hybrid NRPS *trans*-AT-Type I PKS contained the predicted markers. Manual alignment of these
149 two contigs suggested near-identical overlapping sequence (36,638 bases) and, when joined, the
150 merged contig resulted in a 74,672 Kbp BGC (Fig. 1C). The cluster size was in the range of
151 other large *trans*-AT PKS encoding BGCs including pederin (54 Kbp; 41), leinamycin (135.6
152 Kbp; 42) as well as a *cis*-acting AT-PKS, jamaicamide (64.9 Kbp; 43). The combined contigs
153 encompassed what appeared to be a complete BGC that was flanked at the start with a
154 transposase and otherwise unlinked in the assembly to other contiguous DNA. The cluster lacked
155 phylogenetically informative marker genes from which putative taxonomic assignment could be
156 attributed.

157 The antiSMASH results suggested that the BGC appears to be novel with the highest degree
158 of relatedness to pyxipyrrolone A and B (encoded in the *Pyxidicoccus* sp. MCy9557 genome
159 (44), to which only 14% of the genes have a significant BLAST hit to genes in the metagenome-
160 encoded cluster. The ketosynthase (KS) sequences (13 in all) fell into three different sequence
161 groups (40). One was nearly identical (99% amino acid identity) to a previously reported

162 sequence from a targeted KS study of *S. adareanum* microbiome metagenomic DNA (26). The
163 other two were most homologous to KS sequences from *Allochromatium humboldtianum,* and
164 *Dickeya dianthicola* in addition to a number of hypothetical proteins from environmental
165 sequence data sets. The primary polyketide synthase (PKS) component of the BGC (labeled in
166 Fig. 4C, Contig 9) includes 11 of the 13 KS domains which is consistent with a polyketide
167 backbone which has 22 contiguous carbons since each elongation step adds a two-carbon
168 unit. This further supports the hypothesis that this candidate BGC is likely responsible for
169 palmerolide A biosynthesis. A detailed bioinformatic analysis into the step-wise biosynthetic
170 mechanism was conducted by Avalon et al. (40).
171
172 **Taxonomic inference of palmerolide A BGC.** Taxonomic attribution of the BGC was inferred
173 using a real time PCR strategy targeting three coding regions of the putative palmerolide A BGC
174 spanning the length of the cluster (acyltransferase, AT1; hydroxymethylglutaryl Co-A synthase,
175 HCS, and the condensation domain of the non-ribosomal peptide synthase NRPS, Fig. 1C) to
176 assay a *Synoicum* microbiome collection of 63 samples that have been taxonomically classified
177 using Illumina SSU rRNA gene tag sequencing (27). The three gene targets were present in all
178 samples ranging within and between sites at levels from ~$7 \times 10^1 – 8 \times 10^5$ copies per gram of
179 host tissue (Fig. 2A). The three BGC gene targets co-varied across all samples ($r^2 > 0.7$ for all
180 pairs), with the NRPS gene copy levels slightly lower overall (mean: 0.66 and 0.59 copies per ng
181 host tissue for NRPS:AT1 and NRPS:HCS respectively, n=63). We investigated the relationship
182 between BGC gene copies per ng host tissue for each sample and palmerolide A levels
183 determined for the same samples using mass spectrometry, however no correlation was found
184 ($R<0.03$, n=63; 27). We then assessed the semi-quantitative relationship between the occurrence

185 of SSU rRNA amplicon sequence variants (ASV, n=461; 27) and the abundance of the three
186 palmerolide A BGC gene targets. Here, we found a robust correlation (R=0.83 – 0.99) between
187 all 3 gene targets and a single amplicon sequence variant (ASV15) in the core microbiome (45).
188 This ASV is affiliated with the *Opitutaceae* family of the Verrucomicrobium phylum. The
189 *Opitutaceae* family ASV (SaM_ASV15) was a member of the core microbiome as it was
190 detected in 59 of the 63 samples surveyed at varying levels of relative abundance and displayed
191 strong correlations with the abundances of BGC gene targets (Fig. 2B, r^2 =0.68 with AT1, 0.97
192 with HCS, and 0.69 with NRPS, n=63 for all). The only other correlations R > 0.5 were ASVs
193 associated within the "variable" fraction of the microbiome, e.g., one low abundance ASV was
194 present in 24 of 63 samples (45).
195 This result supports the finding of Murray et al. (27) in which gene abundance and natural
196 product chemistry do not reflect a 1:1 ratio in this host-associated system. Neither the semi-
197 quantitative measure of ASV copies nor the real time PCR abundance estimates of the three
198 biosynthetic gene targets correlated with the mass-normalized levels of palmerolide A present in
199 the same samples. As discussed (27), this is likely a result of bioaccumulation in the ascidian
200 tissues. This result provided strong support that the genetic capacity for palmerolide A
201 production was associated with a novel member of the *Opitutaceae*, a taxonomic family with
202 representatives found across diverse host-associated and free-living ecosystems. Although the
203 biosynthetic capacity of this family is not well known (46), recent evidence (36) suggests this
204 family may be a fruitful target for cultivation efforts and natural product surveys.
205
206 **Assembly of the palmerolide BGC-associated *Opitutaceae*-related metagenome assembled**
207 **genome (MAG).** With metagenomes some genomes come together easily – while others present

208 compelling puzzles to solve. Assembly of the *pal* BGC-containing *Opitutaceae* genome was the
209 result of a dedicated effort of binning contigs, gene searches, additional sequencing of samples
210 with high BGC titer, and manual, targeted assembly. Binning efforts with CoAssembly 1 did not
211 result in association of the *pal* BGC with an associated metagenome assembled genome
212 (supplemental materials and methods, Table S3). Therefore, a further round of metagenome
213 sequencing using long read technology (Pacific Biosciences Sequel Systems technology; PacBio)
214 ensued.

215 The 16S rRNA gene ASV occurrence (27) and real time PCR data were used to guide *S.*
216 *adareanum* sample selection for sequencing. Two ascidian samples (Bon-1C-2011 and Del-2B-
217 2011) with high *Opitutaceae* ASV occurrences (ASV_015; > 1000 sequences each – relative
218 abundance of ~ 13.3-15.3 % compared to an overall average of 1.3 ± 2.77% across the 63
219 samples respectively, (27) and high BGC gene target levels (i.e., Bon-1C-2011 (copies per ng ±
220 s.d., n=3): 6.93 ± 1.48 x 10^5 NRPS; 7.52 ± 1.29 x 10^5 AT1; 5.32 ± 1.22 x 10^5 HCS and Del 2B-
221 2011 (copies per ng ± s.d., n=3): 2.04 ± 0.10 x 10^5 NRPS; 2.57 ± 0.41 x 10^5 AT1; 5.84 ± 0.16 x
222 10^5 HCS) were selected for PacBio sequencing. This effort generated 28 GB of data that was
223 used to create a new hybrid CoAssembly 2 which combined all three sequencing technologies.
224 Similar to the assembly with the *Mycale hentscheli*-associated polyketide producers (47), the
225 long-read data set improved the assembly metrics, and subsequent binning resulted in a highly
226 resolved *Opitutaceae*-classified bin (45, supplemental material, Table S3). Interestingly
227 however, the palmerolide BGC contigs still did not cluster with this bin, which we later
228 attributed to binning reliance on sequence depth.

229 We used PacBio circular consensus sequence (CCS) reads to generate and manually edit the
230 assembly for our *Opitutaceae* genome of interest. The resulting 4.3 Mbp genome (Fig. 3A) had a

231 GC content of 58.7% and was resolved into a total of 10 contigs. Five of the contigs (Contigs 1-
232 5) were unique while contigs 6-10 represented highly similar repeated units of the *pal* BGC
233 (labeled *pal* BGC 1, 2, 3, 4, and 5) beginning and ending in linkage gaps. Support for the
234 assembly of a near-complete genome includes manual analysis of the ends of the five unique
235 contigs and underlying read data linking each end of all five contigs to parts of the repeated
236 contigs. The structure of the five repeated contigs was manually evaluated (Fig.1) and while each
237 copy had unique features, all were found to harbor portions of the palmerolide BGC. While the
238 exact placement of each of the five palmerolide contigs could not be positioned with respect to
239 the five unique contigs, the data support a single circular chromosome. Analyses further
240 supporting a complete genome includes the presence of a complete ribosomal operon and
241 estimated CheckM completeness of 96.04 % based on marker genes (presence of which were
242 identified using MetaERG annotation of the *Opitutaceae* genome, Table S3), and 45 tRNA
243 genes.
244 Alignment of all five repeated contigs to the longest palmerolide-containing BGC revealed a
245 long (36,198 base) repeated region that was shared between all 5 contigs with some substantial
246 differences at the beginning of the cluster and only minor differences at the end, indicating 3 full
247 length, and 2 shorter palmerolide BGC-containing contigs (Fig. 1 and 3). This was consistent
248 with coverage estimates based on read-mapping that suggested lower depth at the beginning of
249 the cluster (Fig. 3B). BGC 1 and 3 are nearly identical (over 86,135 bases) with only 2 single
250 nucleotide polymorphisms (SNPs) and an additional 1,468 bases in BGC1 (237 bases at the 5'
251 end and 1231 bases at the 3' end). BGC 4 is 13,470 bases shorter than BGC1 at the 5' end, and 5
252 bases longer than BGC 1 at the 3' end. Alignment of the real time PCR gene targets to the 5 *pal*
253 BGCs provided independent support for the different lengths of the 5 BGCs, as the region

254 targeted by the NRP primers was missing in two of the *pal* BGCs, thus explaining lower NRP:
255 AT or NRP:HCS gene dosages reported above.
256 Interestingly, precedent for naturally occurring multi-copy BGCs to our knowledge, has only
257 recently been found in one other bacterium, an ascidian (*Lissoclinum* sp.)-associated
258 *Opitutaceae*, *Candidatus* Didemnitutus mandela, linked to cytotoxic mandelalide polyketides
259 (36). Likewise, we can invoke a rationale similar to (36) that multiple gene clusters may be
260 linked to biosynthesis of different palmerolide derivatives; see Avalon et al. (40) for retro-
261 biosynthetic predictions of these clusters and annotation of putative enzymatic functions. Gene
262 duplication, loss, and rearrangement processes over evolutionary time, likely explain the source
263 of the multiple copies. At present we do not yet understand the regulatory controls, whether all
264 five are actively transcribed, in situ function and how this may vary amongst host microbiomes.
265
266 ***Phylogenomic characterization of the Opitutaceae-related MAG.*** The taxonomic relationship of
267 the *Opitutacae* MAG to other Verrucomicrobiota was assessed using distance-based analyses
268 with 16S rRNA and average amino acid identity (AAI). Then, it was classified using the GTDB-
269 Tk tool (48), and a phylogenomic analysis based on concatenated ribosomal protein markers.
270 Comparison of 16S rRNA gene sequences amongst other Verrucomicrobia with available
271 genome sequences (that also have 16S rRNA genes; supplemental material, Fig. S1) suggests
272 that the nearest relatives are *Cephaloticoccus primus* CAG34 (similarity of 0.9138), *Optitutus*
273 *terrae* PB90-1 (similarity of 0.9132) and *Geminisphaera coliterminitum* TAV2 (similarity of
274 0.9108). The *Opitutaceae*-affiliated MAG sequence is identical to a sequence (uncultured
275 bacterium clone Tun-3b A3) reported from the same host (*S. adareanum*) in a 2008 study (26);
276 bootstrapping supported a deep branching position in the *Opitutaceae* family. Thus, this

277 Opitutaceae-related MAG appears to be unique – its 16S rRNA gene sequence was not found in
278 bacterioplankton amplicon surveys from the Anvers Island archipelago (n=604 amplicon
279 sequences) and a culture collection reported previously (27), nor were any sequences with
280 identities higher than 95% found following against a large bacterioplankton amplicon data set
281 from further north in the Antarctic Peninsula (32,941 sequence clusters derived from 44.6
282 million sequences; NCBI Bioproject PRJNA316748). Likewise, when searched against the
283 global nr GenBank database representing environmental data sets the highest sequence identities
284 found so far are < 96% identity over 88% of the complete sequence.

285 When characterizing the MAG using AAI metrics (average nucleotide identity, ANI, found
286 no closely related genomes) the closest genomes were environmental metagenome assemblies
287 from the South Atlantic TOBG_SAT_155 (53.08 % AAI) and WB6_3A_236 (52.71 % AAI);
288 and the two closest isolate type genomes were *Nibricoccus aquaticus* str. NZ CP023344 (52.82
289 % AAI) and *Opitutus terrae* str. PB90 (52.75 % AAI). The Microbial Genome Atlas (MiGA)
290 support for the MAG belonging in the *Opitutaceae* family was weak (p-values of 0.5). Attempts
291 to classify this MAG using GTDB-Tk (47) were hampered by the fact we have no real
292 representative in the genome databases, resulting in low confidence predictions at the species or
293 genus levels (see the supplemental material for details).

294 Verrucomicrobia exhibit free-living and host-associated lifestyles in a multitude of terrestrial
295 and marine habitats on Earth. We performed a meta-analysis of Verrucomicrobia genomes, with
296 an emphasis on marine and host-associated *Opitutaceae*, to establish more confidence in the
297 phylogenetic position of the *Opitutaceae* MAG. The analysis was based on 24 conserved
298 proteins – 21 ribosomal proteins and three additional conserved proteins (InfB, lepA, pheS). The
299 diversity of the *Opitutaceae* family, and of Verrucomicrobia in general, is largely known from

300 uncultivated organisms in which there are 20 genera in GTDB (release 05-RS95), 2 additional
301 genera in the NCBI taxonomy database, and numerous unclassified single amplified genomes
302 (SAGs); in all, only eight genera have cultivated representatives. Given the uneven
303 representations of the 24 proteins across all (115) genomes assessed (MAGs and SAGs are often
304 incomplete), we selected a balance of 16 proteins across 48 genomes to assess phylogenomic
305 relatedness across the *Opitutaceae* (Fig. 4). Here too, as seen with the 16S rRNA gene
306 phylogenetic tree, the *S. adareanum*-*Opitutaceae* MAG held a basal position compared to the
307 other *Opitutaceae* genomes in the analysis.
308
309 ***Ca.* S. palmerolidicus relative abundance estimates and ecological inference.** The relative
310 abundance of *Opitutaceae* bin 8 was estimated in the shotgun metagenomic samples by mapping
311 the NGS reads back to the assembled MAG across the four *S. adareanum* samples collected.
312 This indicated varying levels of genome coverage in the natural samples, with the two samples
313 selected based on real time PCR-quantified high BGC copy number being clearly enriched in this
314 strain (44.70 % of reads mapped to Bon-1C-2011 and 36.78 % to Del-2B-2011, Table 1). These
315 levels are higher than estimates of relative abundance derived from the 16S rRNA gene amplicon
316 surveys (estimated at 13.33 and 15.34 % respectively) for the same samples. This is likely a
317 result of the single-copy nature of the ribosomal operon in *Opitutaceae* bin 8 vs. other taxa with
318 multiple rRNA operon copies that could thus be over-represented in the core microbiome library
319 (e.g., *Pseudovibrio* sp. str. PSC04-5.I4 has 9 and *Microbulbifer* sp. is estimated at 4.1 ± 0.8
320 based on 9 finished *Microbulbifer* genomes available at the Integrated Microbial Genomes
321 Database). All host *S. adareanum* lobes surveyed (n=63) in the Anvers Island regional survey
322 contained high levels (0.49 – 4.06 mg palmerolide A x g^{-1} host dry weight) of palmerolide A

323 (27), and variable, yet highly concordant levels of the *pal* BGCs and 16S rRNA ASV levels (Fig.
324 2). Despite the natural population structure sampled here (four single host lobes), the bin-level
325 sequence variation was low (ranging from 72-243 SNPs) when the PacBio reads were mapped
326 back to the Opitutaceae bin 8 (Table 1). This suggests maintenance of a relatively invariant
327 population at the spatial and temporal scales of this coastal Antarctic region while highlighting
328 our limited understanding of the biogeographical extent of the *S. adareanum*-symbiont-
329 palmerolide relationship across a larger region of the Southern Ocean.
330 Several questions remain with regard to the in situ function of palmerolide A (a eukaryotic
331 V-ATPase inhibitor in human cell line assays (28) in this cryohabitat: how and why is it
332 bioaccumulated by the host? Overall, the study of natural products in high latitude marine
333 ecosystems is in its infancy. This palmerolide producing, ascidian-associated, *Opitutaceae*
334 provides the first Antarctic example in which a well-characterized natural product has been
335 linked to the genetic information responsible for its biosynthesis. Gaining an understanding of
336 environmental and biosynthetic regulatory controls, establishing integrated transcriptomic,
337 proteomic, and secondary metabolome expression in the environment will also reveal whether
338 the different clusters are expressed in situ. In addition to ecological pursuits, the path to clinical
339 studies of palmerolide will require genetic or cultivation efforts. At present, we hypothesize that
340 cultivation of *Opitutaceae* bin 8 may be possible, given the lack of genome reduction or of other
341 direct evidence for host-associated dependencies (e.g., a number of central carbohydrate and
342 energy metabolism pathways appear to be present).
343
344 ***Candidatus* Synoicihabitans palmerolidicus genome attributes**. The Antarctic ascidian,
345 *Synoicum adareanum*, harbors a dense community of bacteria that has a conserved core set of

346 taxa (27). The near complete ~4.30 Mbp *Opitutaceae* bin 8 metagenome assembled genome (Fig.
347 3) represents one of the core members. This MAG is remarkable in that it encodes for five 36-74
348 Kbp copies of the candidate BGCs that are implicated in biosynthesis of palmerolide A and
349 possibly other palmerolide compounds. Intriguingly, this genome does not seem to show
350 evidence of genome reduction as found in *Candidatus* Didemnitutus mandela (36); the other
351 ascidian-associated *Opitutaceae* genome currently known to encode multiple BGC gene copies.
352 This is the first *Opitutaceae* genome characterized from a permanently cold, ~ -1.8 - 2 °C, often
353 ice-covered ocean ecosystem. This genome encodes one rRNA operon, 45 tRNA genes, and an
354 estimated 5058 coding sequences. Based on the low (< 92%) SSU rRNA gene identity and low
355 (< 54% AAI) values to other genera in the *Opitutaceae*, along with the phylogenomic position of
356 the *Opitutaceae* bin 8, the provisional name "*Candidatus* Synoicihabitans palmerolidicus" (*Ca*.
357 S. palmerolidicus) is proposed for this novel verrucomicrobium. The genus name
358 *Synoicihabitans* (Syn.o.i.ci.ha'bitans. N.L. neut. N. *Synoicum* a genus of ascidians; L. pres. part
359 *habitans* inhabiting; N.L. masc. n.) references this organism as an inhabitant of the ascidian
360 genus *Synoicum*. The species name *palmerolidicus* (pal.me.ro.li'di.cus. N.L. neut.
361 n. *palmerolidum* palmerolide; N.L. masc. adj.) designates the species as pertaining to
362 palmerolide.
363 The GC content of 58.7% is rather high compared to other marine *Opitutaceae* genomes
364 (ave. 51.49 s.d. 0.02, n=12), yet is ~ average for the family overall (61.58 s.d. 0.06, n=69;
365 supplemental material, Table S4). MetaERG includes metagenome assembled genomes available
366 in the GTDB as a resource for its custom GenomeDB that new genomes are annotated against.
367 This was a clear advantage in annotating the *Ca.* S. palmerolidicus genome as Verrucomicrobia
368 genomes are widely represented by uncultivated taxa. Likewise, antiSMASH was an invaluable

369 tool for *pal* BGC identification and domain structure annotation. This formed the basis to derive
370 a predicted step-wise mechanism of *pal* biosynthesis (40).
371
372 ***Ca*. S. palmerolidicus genome structure, function and host-associated features.** Beyond the
373 *pal* BGCs, the *Ca*. S. palmerolidicus genome encodes a variety of additional interesting
374 structural and functional features that provide insight its lifestyle. Here we provide but a brief
375 synopsis. In addition to the repeated BGCs, three additional repeated elements with two nearly
376 identical copies each (15.3 Kbp, 17.0 Kbp, 27.4 Kbp) were identified during the assembly
377 process (Fig. 3A). These coded for 20, 25 and 41 CDSs respectively, were in some cases flanked
378 by transposase/integrases (both internal and proximal) and had widespread homology with
379 Verrucomicrobia orthologs. The contents of the three repetitive elements were not shared
380 amongst each other.
381 Annotations were assigned to a little more than half of the CDSs in the 15.3 Mbp repeat
382 which is predicted to encode for xylose transport, two sulfatases, two endonucleases and a
383 MacB-like (potential macrolide export) periplasmic core protein. Xylan might be sourced from
384 seaweeds (49) or even the ascidian, as it is a minor component of the tunic cellulose (50).
385 Related to this, an endo-1,4-beta-xylanase which has exoenzyme activity in some
386 microorganisms (51) was identified elsewhere in the genome suggesting the potential for xylose
387 metabolism. Altogether, eight sulfatase copies were identified in this host-associated genome
388 (four in the 15.5 Mbp repeat elements). These may be involved in catabolic activities of
389 sulfonated polysaccharides, and possibly as *trans*-acting elements in palmerolide biosynthesis
390 (40). In addition to the MacB-like CDSs found in this repeat, 13 other MacB-homologs were
391 present in the genome – none of which were associated with the *pal* BGCs (supplemental

392 material, Fig. S2a). MacB is a primary component of the macrolide tripartite efflux pump that
393 operates as a mechanotransmission system involved both in antibiotic resistance and antibiotic
394 export depending on the size of the macrolide molecule (52). However, two additional elements
395 required for this pump to be functional, an intramembrane MacA and an outer membrane protein
396 TolC, were not co-located elsewhere in the genome. MacA may be missing, as searches against
397 the of *Ca*. S. palmerolidicus genome with to two other verrucomicrobia-associated MacA CDSs
398 were not identified using BLAST queries (Peat Soil MAG SbV1 SBV1_730043 and *Ca*.
399 Udaeobacter copiosis KAF5408997.1; (53). At least nine MacB CDSs were flanked by a FstX-
400 like permease family protein; the genomic structures of which were quite complex including
401 several with multiple repeated domains. Detailed transporter modeling is beyond the scope of
402 this work, but it is likely that these proteins are involved in signaling of cell division machinery
403 rather than macrolide transport (54).
404 Predicted CDSs in the 17.0 Mbp repeat included sugar binding and transport domains, as
405 well as domains encoding rhamnosidase, arabinofuranosidase, and other carbohydrate catabolism
406 functions. About half the proteins encoded in the 27.4 Mbp repeat were unknown in function,
407 and of the remaining characterized suggested diverse potential functional capacities. For
408 example, a zinc carboxypeptidase (1 of 3 in the genome), multidrug and toxic compound
409 transporter (MatE/NorM), and an exodeoxyribonuclease were identified.
410 The *Ca.* S. palmerolidicus genome has a number of features that suggest it is adapted to a
411 host-associated lifestyle, and several of these features were reported recently for two related
412 sponge-associated Opitutales metagenome bins (*Petrosia ficiformis*-associated bins 0 and 01,
413 Fig. 4; (55). These include identification of a bacterial microcompartment (BMC) 'super locus'.
414 Such loci were recently reported to be enriched in host-associated Opitutales genomes when

415 compared to free-living relatives. The structural proteins for the BMC were present as were other
416 conserved Planctomyces-Verrucomicrobia BMC genes (56). As in the sponge *Pectoria ficiformis*
417 metagenome bins, enzymes for carbohydrate (rhamnose) catabolism and modification were
418 found adjacent to the BMC locus (supplemental material, Fig. S3), in addition to the two that
419 were found in the 27.4 Mbp repeat. The genome did not appear to encode the full complement of
420 enzymes required for fucose metabolism, though a few alpha-L-fucosidases were identified.
421 Further evidence for carbohydrate metabolism was supported through genome similarity
422 searches with the CAZY database (57), including 7 identified carbohydrate binding modules, a
423 carbohydrate esterase, 14 glycoside hydrolases, 6 glycosyl transferases and a polysaccharide
424 lyase. In addition, three bacterial cellulases (PF00150, cellulase family A; glycosyl hydrolase
425 family 5) were identified, all with the canonical conserved glutamic acid residue. These appear to
426 have different evolutionary histories in which each variant has nearest neighbors in different
427 bacterial phyla (supplemental material, Fig. S2b) matching between 68% identity for Protein
428 J6386 03765 to *Lacunisphaera limnophila*, 57.5% identity for Protein J6386 22340 with a
429 cellulase from a shipworm symbiont *Alteromonadaceae* (*Terridinibacter* sp.), and 37.5%
430 sequence identity to a Bacteroidetes bacterium. This suggests the potential for cellulose
431 degradation – which is consistent with ascidians being the only animals known to produce
432 cellulose for its skeletal structure (50). In addition to the BGCs, the enzymatic resources in this
433 genome (e.g., xylan and cellulose hydrolysis) are a treasure trove rich with biotechnological
434 potential. Support for a Type II secretion system (i.e. GspD, GspE, GspG and GspO), common to
435 gram-negative bacteria which secrete folded proteins (e.g. hydrolytic enzymes required for

436 survival in host environments) from the periplasm into the extracellular environment (58) were
437 detected in the *Ca*. S. palmerolidicus genome.
438 Chemotaxis and flagellar biosynthesis are factors required for horizontally acquired
439 symbionts, as has been established with the *Vibrio fischeri-Euprymna scolopes* symbiosis (59).
440 The *Ca*. S. palmerolidicus genome encodes a chemotaxis system (e.g. CheA, CheW, CheR,
441 CheB, Che C, methyl-accepting chemotaxis domains) with flagellar motors in addition to a
442 number of other elements of flagellar biosynthesis, which is consistent with horizontal
443 acquisition by the host. The methyl-accepting chemotaxis system in *V. fischeri* responds to
444 chitobiose production by *E. scolopes* (59) – it will be interesting to further unravel the details of
445 Ca. S. palmerolidicus' cellular biology to understand the chemotaxis stimulants, preferential
446 cellular localization of palmerolide production and resistance mechanisms of the host to the
447 potent vacuolar ATPase, as well as products made by others in the *S. adareanum* microbiome.
448 Other indicators of host-association and palmerolide production include T-A domains,
449 multidrug exporters, and the potential for palmerolide transport and cofactor biosynthesis. T-A
450 domains were also prevalent in the *Petrosia ficiformis*-associated bins 0 and 01 (55). The *Ca*. S.
451 palmerolidicus genome encoded at least 22 TA-related genes including multiple MazG and
452 AbiEii toxin type IV TA systems, AbiEii-Phd_YefM type II toxin-antitoxin systems, along with
453 genes coding for PIN domains, Zeta toxin, RelB, HipA, MazE and MraZ. In addition to the MatE
454 (found in the 27.4 Mbp repeat) two other multidrug export systems with homology to MexB and
455 MdtB were identified. This analysis also resulted in identifying a putative AbiEii toxin
456 (PF13304) with homology to SyrD, a cyclic peptide ABC type transporter that was present in all
457 5 BGCs (Fig. 1C; with 52.7% BLAST percent identity to a *Desulfamplus* sp. homolog over the
458 full length of the protein, and a variety of other bacteria including an *Opitutaceae*-related strain

459 with similar levels of identity). This transporter is encoded downstream of the large polyketide
460 gene clusters following the acyl transferase domains and precedes the predicted *trans*-acting
461 domains at the 3' end of the BGC. Given its genomic position, this protein is a candidate for
462 palmerolide transport. The *Ca*. S. palmerolidicus genome also encodes the potential for
463 pantothenate biosynthesis via ilvD, ilvE, panB, ilvC, panD, and panC (60), which is consistent
464 with palmerolide biosynthesis in which the 4'-phosphopantetheinyl prosthetic group interacts
465 with acyl carrier proteins for multimodular assembly. Some symbionts, e.g. *Ca*. Entotheonella
466 sp., are auxotrophic for pantothenate and likely acquire it from other microbiome members (61).
467 Unlike in *Ca*. D. mandela (36), there does not appear to be ongoing genome reduction, which
468 may suggest that the *S. adareanum*-*Ca*. S palmerolidicus relationship is more recent, and/or that
469 the relationship is commensal rather than interdependent. Likewise, we suspect that the
470 pseudogene content may be high as several CDS appear to be truncated, in which redundant CDS
471 of varying lengths were found in several cases (including the MacB). There is evidence of lateral
472 gene transfer acquisitions of cellulase and numerous other enzymes that may confer ecological
473 advantages through the evolution of this genome. Similarly, the origin of the *pal* BGCs and how
474 recombination events play out in the success of this Antarctic host-associated system in terms of
475 adaptive evolution (62), not to mention the ecology of *S. adareanum* is a curiosity. This phylum
476 promises to be an interesting target for further culture-based and cultivation-free studies –
477 particularly in the marine environment, in which they have been less well studied compared to
478 terrestrial free-living and host associated systems.
479 Together, it appears that the genome of *Ca*. S. palmerolidicus is equipped for life in this host-
480 associated interactive ecosystem that stands to be one of the first high latitude marine
481 invertebrate-associated microbiomes with a genome-level understanding – and one that produces

482 a highly potent natural product, palmerolide A. This system holds promise for future research
483 now that we have identified the producing organism and *pal* BGC. We still have much to learn
484 about the ecological role of palmerolide A – if is it involved predation avoidance, antifouling,
485 antimicrobial defense or some other yet to be recognized aspect of life in the frigid, often ice-
486 covered and seasonally light-limited waters of the Southern Ocean.
487

488 **Materials and Methods**

489 **Sample Collection.** *S. adareanum* lobes were collected in the coastal waters off Anvers Island,
490 Antarctica and stored at -80 °C until processing (supplemental material, Table S1). See the
491 supplemental material and (27) for details of sample collection, microbial cell preparation and
492 DNA extraction.
493

494 **Metagenome sequencing.** Three rounds of metagenome sequencing were conducted, the details
495 of which are in the supplemental material. This included an initial 454 pyrosequencing effort
496 with a bacterial-enriched metagenomic DNA preparation from *S. adareanum* lobe (Nor2c-2007).
497 Next, an Ion Proton System was used to sequence a metagenomic DNA sample prepared from *S.*
498 *adareanum* lobe Nor2a-2007. Then two additional *S. adareanum* metagenome DNA samples
499 (Bon-1C-2011 and Del-2b-2011) selected based on high copy numbers of the palmerolide A
500 BGC (see real time PCR Methods, supplemental material) and sequenced using Pacific
501 Biosciences Sequel Systems technology.
502

503 **Metagenome assembly, annotation and binning**. Raw 454 metagenomic reads (1,570,137
504 single end reads, 904,455,285 bases) were assembled by Newbler (63) v2.9 (Life Technologies,

505 Carlsbad, CA, flags: -large -rip -mi 98 -ml 80), while Ion Proton metagenomic reads (89,330,870
506 reads, 17,053,251,055 bases) were assembled using SPAdes (64) v3.5 (flags: --iontorrent). Both
507 assembled datasets were merged with MeGAMerge (65) v1.2 and produced 86,387 contigs with
508 a maximum contig size 153,680 and total contig size 144,953,904 bases (CoAssembly 1). To
509 achieve more complete metagenome coverage and facilitate metagenome assembled genome
510 assembly, a Circular Consensus Sequence (CCS) protocol (PacBio) was used to obtain high
511 quality long reads on two samples Bon-1C-2011 and Del-2b-2011. The 5,514,426 PacBio reads
512 were assembled with aforementioned assembled contigs (CoAssembly 1) on EDGE
513 Bioinformatics using wtdbg2 (66), a fast and accurate long-read assembler. The contigs were
514 polished with three rounds of polishing by Racon (67) into a second Coassembly (CoAssembly
515 2) which has 4,215 contigs with a maximum contig size 2,235,039 and total size 97,970,181
516 bases. Lastly, a manual approach was implemented to arrive at assembly of the MAG of interest
517 the details of which are described in the supplemental material.
518 The contigs from both co-assemblies 1 and 2 were submitted initially to the EDGE
519 bioinformatics platform (68) for sequence annotation using Prokka (69) v1.13 and taxonomy
520 classification using BWA (70) mapping to NCBI RefSeq (Version: NCBI 2017 Oct 3).
521 Bioinformatic predictions of natural product potential was performed using the antibiotics and
522 secondary metabolite analysis shell (antiSMASH, bacterial versions 3.0, 4.0 and 5.0 (38, 39, 71).
523 This tool executed contig identification, annotation and analysis of secondary metabolite
524 biosynthesis gene clusters on both CoAssemblies 1 and 2 (> 1 Kbp and > 40 Kbp data sets). As
525 most of our attention was focused on analysis the *Ca*. S. palmerolidicus assembled metagenome,
526 we also used MetaERG (72) as the primary pipeline for metagenome annotation of the ten final

527 contigs in addition to NCBI's PGAP pipeline (see (45) for GFF dataset). There were 5186 coding
528 sequences predicted in the MetaERG annotation and 5186 in NCBI's PGAP annotation.
529 MaxBin (73) and MaxBin2 (74) were used to form metagenome bins for both CoAssembly 1
530 and 2. CheckM v1.1.11 (75) and v.1.1.12, and GTDB-Tk v.1.0.2 (48) were used to verify bin
531 quality and taxonomic classification. See supplemental material for details. In order to assess the
532 representation of assembled *Opitutaceae* genome across the 4 environmental samples used for
533 metagenome sequencing (resulting from MaxBin2 binning of CoAssembly 2), we used BWA to
534 map the CCS reads to each metagenome data set.
535
536 **Real time PCR.** Gene targets (non-ribosomal peptide synthase, acyltransferase, and 3-
537 hydroxymethylglutaryl coenzyme A synthase) were selected at different positions along the
538 length of the candidate BGC. Supplemental material, Table S5 lists the primer and the GBlocks
539 synthetic positive control sequence. Metagenomic DNA extracts from a large *S. adareanum*
540 sample set (n=63 *S. adareanum* lobes from 21 colonies), all containing high levels of
541 palmerolide A (27), were screened with the real time PCR assays on a Quant Studio 3 (Thermo
542 Fisher Scientific, Inc.; see supplemental material text for details of controls and analysis).
543
544 **Phylogenomic analyses.** A phylogenomic analysis of the assembled *Opitutaceae* MAG was
545 conducted based on shared ribosomal RNA and ribosomal proteins amongst 46 and 48 reference
546 genomes respectively, out of 115 genomes in total, mined from various databases (NCBI, GTDB
547 and IMG) for uncultivated and cultivated microorganisms identified in the Verrucomicrobia
548 phylum (supplemental material, Table S6). The details of these analyses are described in the
549 supplemental material. In addition, we used MiGA (NCBI Prokaryotic taxonomy and the

550 environmental TARA Oceans (Tully) databases; accessed August 2020) and GTDB-Tk (ver.
551 1.3.0) tools for MAG taxonomic classification.
552 Phylogenetic analysis of the MacB CDS sequences were retrieved from MetatERG annotated
553 *Ca*. S. palmerolidicus contigs and homologs were retrieved from the NCBI based on BLAST
554 results. Maximum likelihood analysis was conducted on 994 aligned (MUSCLE) positions using
555 RAxML v.8.2.12 using the PROTGAMMALG model and 550 bootstrap replicates. For the
556 phylogenetic analysis of the cellulase CDS, homologs were retrieved from the NCBI based on
557 BLAST results resulting in 19 sequences and 496 aligned positions (ClustalOmega) was also
558 conducted using RAxML v.8.2.12 under the PROTGAMMALG model of evolution with 1000
559 bootstraps.
560
561 **Data Availability**
562 The data (biosamples, SRA depositions) associated with this study is associated with NCBI
563 BioProject PRJNA662631. The *Ca.* S. palmerolidicus biosample ID is SAMN18473105 and
564 genome accession JAGGDC000000000. An annotation data set (.gff) for the Ca. S.
565 palmerolidicus MAG resulting from the metaERG pipeline is available (45).
566
567 **Acknowledgments**
568 Support for this research was provided in part by the National Institute of Health award
569 (CA205932) to A.E.M., B.J.B., and P.S.G.C., with additional support from National Science
570 Foundation awards (OPP-0442857, ANT-0838776, and PLR-1341339) to B.J.B., and Desert
571 Research Institute (Institute Project Assignment) to A.E.M..

572 The assistance of several collaborators and students are acknowledged including C. Amsler,
573 M. Amsler, L. Bishop, J. Cuce, B. Dent, N. Ernster, C. Gleasner, A. Maschek, A. Shilling, L.
574 Siao, S. Thomas, and the Palmer Station science support staff. We especially would like to
575 acknowledge the help of A. Oren in assisting with bacterial nomenclature proposed here.
576 Likewise, we thank J.T. Hollibaugh and A.L. Reysenbach for comments on previous drafts of the
577 manuscript.

578

579 **References**

580 1. Simmons TL, Coates RC, Clark BR, Engene N, Gonzalez D, Esquenazi E, Dorrestein PC,
581 Gerwick WH. 2008. Biosynthetic origin of natural products isolated from marine
582 microorganism-invertebrate assemblages. Proc Natl Acad Sci USA 105:4587-4594.
583 2. Morita M, Schmidt EW. 2018. Parallel lives of symbionts and hosts: chemical mutualism
584 in marine animals. Nat Prod Rep 35:357-378.
585 3. Piel J, Hui DQ, Fusetani N, Matsunaga S. 2004. Targeting modular polyketide synthases
586 with iteratively acting acyltransferases from metagenomes of uncultured bacterial
587 consortia. Environ Microbiol 6:921-927.
588 4. Schmidt EW, Nelson JT, Rasko DA, Sudek S, Eisen JA, Haygood MG, Ravel J. 2005.
589 Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the
590 cyanobacterial symbiont of *Lissoclinum patella*. Proc Natl Acad Sci USA 102:7315-
591 7320.
592 5. Piel J. 2009. Metabolites from symbiotic bacteria. Natural Product Reports 26:338-362.
593 6. Sunagawa S, Woodley CM, Medina M. 2010. Threatened corals provide underexplored
594 microbial habitats. Plos One 5:e9554.
595 7. Behrendt L, Larkum AWD, Trampe E, Norman A, Sorensen SJ, Kuhl M. 2012. Microbial
596 diversity of biofilm communities in microniches associated with the didemnid ascidian
597 Lissoclinum patella. ISME J 6:1222-1237.
598 8. Webster NS, Taylor MW. 2012. Marine sponges and their microbial symbionts: love and
599 other relationships. Environ Microbiol 14:335-346.
600 9. McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Loso T, Douglas AE,
601 Dubilier N, Eberl G, Fukami T, Gilbert SF, Hentschel U, King N, Kjelleberg S, Knoll
602 AH, Kremer N, Mazmanian SK, Metcalf JL, Nealson K, Pierce NE, Rawls JF, Reid A,
603 Ruby EG, Rumpho M, Sanders JG, Tautz D, Wernegreen JJ. 2013. Animals in a bacterial
604 world, a new imperative for the life sciences. Proc Natl Acad Sci USA 110:3229-3236.
605 10. Gray J. 2001. Antarcic marine benthic biodiversity in a world-wide latitudinal context.
606 Polar Biol 24:633-641.
607 11. Clarke A. 2008. Antarctic marine benthic diversity: patterns and processes. J Exp Mar
608 Biol Ecol 366:48-55.

12. Piepenburg D, Archambault P, Ambrose WG, Jr., Blanchard AL, Bluhm BA, Carroll ML, Conlan KE, Cusson M, Feder HM, Grebmeier JM, Jewett SC, Levesque M, Petryashev VV, Sejr MK, Sirenko BI, Wlodarska-Kowalczuk M. 2011. Towards a pan-Arctic inventory of the species diversity of the macro- and megabenthic fauna of the Arctic shelf seas. Mar Biodivers 41:51-70.
13. von Salm J, Schoenrock K, McClintock J, Amsler C, Baker B. 2018. The status of marine chemical ecology in Antarctica: form and function of unique high-latitude chemistry, p 69, Chemical Ecology: The ecological impacts of marine natural products, 1st ed doi:10.1201/9780429453465. CRC Press, Boca Raton.
14. McClintock JB, Amsler CD, Baker BJ. 2001. Introduction to the symposium: Antarctic marine biology. Am Zool 41:1-2.
15. Bakus GJ, Green G. 1974. Toxicity in sponges and holothurians - geographic pattern. Science 185:951-953.
16. Soldatou S, Baker BJ. 2017. Cold-water marine natural products, 2006 to 2016. Nat Prod Rep 34:585-626.
17. Avila C, Taboada S, Nunez-Pons L. 2008. Antarctic marine chemical ecology: what is next? Mar Ecol-Evol Persp 29:1-71.
18. Webster NS, Negri AP, Munro M, Battershill CN. 2004. Diverse microbial communities inhabit Antarctic sponges. Environ Microbiol 6:288-300.
19. Webster NS, Bourne D. 2007. Bacterial community structure associated with the Antarctic soft coral, *Alcyonium antarcticum*. FEMS Microbiol Ecol 59:81-94.
20. Murray AE, Rack FR, Zook R, Williams MJM, Higham ML, Broe M, Kaufmann RS, Daly M. 2016. Microbiome composition and diversity of the ice-dwelling sea anemone, *Edwardsiella andrillae*. Integr Comp Biol 56:542-555.
21. Rodriguez-Marconi S, De la Iglesia R, Diez B, Fonseca CA, Hajdu E, Trefault N. 2015. Characterization of bacterial, archaeal and eukaryote symbionts from Antarctic sponges reveals a high diversity at a three-domain level and a particular signature for this ecosystem. Plos One 10:e0138837.
22. Godinho VM, de Paula MTR, Silva DAS, Paresque K, Martins AP, Colepicolo P, Rosa CA, Rosa LH. 2019. Diversity and distribution of hidden cultivable fungi associated with marine animals of Antarctica. Fungal Biol 123:507-516.
23. Steinert G, Wemheuer B, Janussen D, Erpenbeck D, Daniel R, Simon M, Brinkhoff T, Schupp PJ. 2019. Prokaryotic diversity and community patterns in Antarctic continental shelf sponges. Front Mar Sci 6:297.
24. Sacristan-Soriano O, Criado NP, Avila C. 2020. Host species determines symbiotic community composition in Antarctic sponges (Porifera: *Demospongiae*). Front Mar Sci 7:474.
25. Moreno-Pino M, Cristi A, Gillooly JF, Trefault N. 2020. Characterizing the microbiomes of Antarctic sponges: a functional metagenomic approach. Sci Rep UK 10:645.
26. Riesenfeld CS, Murray AE, Baker BJ. 2008. Characterization of the microbial community and polyketide biosynthetic potential in the palmerolide-producing tunicate, *Synoicum adareanum*. J Nat Prod 71:1812-1818.
27. Murray AE, Avalon NE, Bishop L, Davenport KW, Delage E, Dichosa AEK, Eveillard D, Higham ML, Kokkaliari S, Lo C-C, Riesenfeld CS, Young RM, Chain PSG, Baker BJ. 2020. Uncovering the core microbiome and distributions of palmerolide in *Synoicum adareanum* across the Anvers Island archipelago, Antarctica. Mar Drugs 18:298.

655 28. Diyabalanage T, Amsler CD, McClintock JB, Baker BJ. 2006. Palmerolide A, a cytotoxic
656 macrolide from the Antarctic tunicate *Synoicum adareanum*. J Am Chem Soc 128:5630-
657 5631.
658 29. Franco LH, Joffe EBD, Puricelli L, Tatian M, Seldes AM, Palermo JA. 1998. Indole
659 alkaloids from the tunicate *Aplidium meridianum*. J Nat Prod61:1130-1132.
660 30. Miyata Y, Diyabalanage T, Amsler CD, McClintock JB, Valeriote FA, Baker BJ. 2007.
661 Ecdysteroids from the antarctic tunicate *Synoicum adareanum*. J Nat Prod 70:1859-1864.
662 31. Seldes AM, Brasco MFR, Franco LH, Palermo JA. 2007. Identification of two
663 meridianins from the crude extract of the tunicate *Aplidium meridianum* by tandem mass
664 spectrometry. Nat Prod Rep21:555–563.
665 32. Chen L, Hu JS, Xu JL, Shao CL, Wang GY. 2018. Biological and chemical diversity of
666 ascidian-associated microorganisms. Mar Drugs 16:362.
667 33. Dou X, Dong B. 2019. Origins and Bioactivities of Natural Compounds Derived from
668 Marine Ascidians and Their Symbionts. Mar Drugs 17:670.
669 34. Rath CM, Janto B, Earl J, Ahmed A, Hu FZ, Hiller L, Dahlgren M, Kreft R, Yu FA,
670 Wolff JJ, Kweon HK, Christiansen MA, Hakansson K, Williams RM, Ehrlich GD,
671 Sherman DH. 2011. Meta-omic characterization of the marine invertebrate microbial
672 consortium that produces the chemotherapeutic natural product ET-743. ACS Chem Biol
673 6:1244-1256.
674 35. Kwan JC, Donia MS, Han AW, Hirose E, Haygood MG, Schmidt EW. 2012. Genome
675 streamlining and chemical defense in a coral reef symbiosis. Proc Natl Acad Sci USA
676 109:20655-20660.
677 36. Lopera J, Miller IJ, McPhail KL, Kwan JC. 2017. Increased biosynthetic gene dosage in a
678 genome-reduced defensive bacterial symbiont. mSystems 2:e00096-17.
679 37. Shenkar N, Swalla BJ. 2011. Global diversity of Ascidiacea. PLoS One 6:e20657.
680 38. Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA,
681 Muller R, Wohlleben W, Breitling R, Takano E, Medema MH. 2015. antiSMASH 3.0-a
682 comprehensive resource for the genome mining of biosynthetic gene clusters. Nuc Acids
683 Res 43:W237-W243.
684 39. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T.
685 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nuc
686 Acids Res 47:W81-W87.
687 40. Avalon N, Murray A, Daligault H, Lo C, Davenport K, Dichosa A, Chain P, Baker B.
688 Bioinformatic and mechanistic analysis of the palmerolide PKS-NRPS biosynthetic
689 pathway from the microbiome of an Antarctic ascidian.
690 bioRxiv:doi.org/10.1101/2021.04.05.438531.
691 41. Piel J. 2002. A polyketide synthase-peptide synthetase gene cluster from an uncultured
692 bacterial symbiont of *Paederus* beetles. Proc Natl Acad Sci USA 99:14002-7.
693 42. Cheng YQ, Tang GL, Shen B. 2002. Identification and localization of the gene cluster
694 encoding biosynthesis of the antitumor macrolactam leinamycin in *Streptomyces*
695 *atroolivaceus* S-140. J Bacteriol 184:7013-24.
696 43. Edwards DJ, Marquez BL, Nogle LM, McPhail K, Goeger DE, Roberts MA, Gerwick
697 WH. 2004. Structure and biosynthesis of the jamaicamides, new mixed polyketide-
698 peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*. Chem Biol
699 11:817-33.

700 44. Kjaerulff L, Raju R, Panter F, Scheid U, Garcia R, Herrmann J, Muller R. 2017.
701 Pyxipyrrolones: structure elucidation and biosynthesis of cytotoxic myxobacterial
702 metabolites. Angew Chem Int E. Engl 56:9614-9618.
703 45. Murray, AE, Lo, C-C, Daligault, HE, Avalon, NE, Read, RW, Davenport, KW, Higham,
704 ML, Kunde, Y, Dichosa, AEK, Baker, BJ, Chain, PSG 2021. Supplementary information
705 provided with Murray et al. "Discovery of an Antarctic ascidian-associated uncultivated
706 Verrucomicrobia with antimelanoma palmerolide biosynthetic potential.", Dryad,
707 Dataset, https://doi.org/10.5061/dryad.8sf7m0cpp.
708 46. Blin K, Andreu VP, de los Santos ELC, Del Carratore F, Lee SY, Medema MH, Weber
709 T. 2019. The antiSMASH database version 2: a comprehensive resource on secondary
710 metabolite biosynthetic gene clusters. Nuc Acids Res 47:D625-D630.
711 47. Storey MA, Andreassend SK, Bracegirdle J, Brown A, Keyzers RA, Ackerley DF,
712 Northcote PT, Owen JG. 2020. Metagenomic exploration of the marine sponge *Mycale*
713 *hentscheli* uncovers multiple polyketide-producing bacterial symbionts. Mbio 11:e02997-
714 19.
715 48. Chaumeil P-A, Mussig A, Hugenholtz P, Parks D. 2020. GTDB-Tk: a toolkit to classify
716 genomes with the Genome Taxonomy Database. Bioinformatics 36:1925–1927.
717 49. Qeshmi FI, Homaei A, Fernandes P, Hemmati R, Dijkstra BW, Khajeh K. 2020.
718 Xylanases from marine microorganisms: A brief overview on scope, sources, features
719 and potential applications. BBA-Proteins Proteom 1868:140312.
720 50. Zhao Y, Li J. 2014. Excellent chemical and material cellulose from tunicates: diversity in
721 cellulose production yield and chemical and morphological structures from differnet
722 tunicate species. Cellul Chem Technol 21:3427-3441.
723 51. Juturu V, Wu JC. 2014. Microbial exo-xylanases: a mini review. Appl Biochem
724 Biotechnol 174:81-92.
725 52. Greene NP, Kaplan E, Crow A, Koronakis V. 2018. Antibiotic resistance mediated by the
726 MacB ABC transporter family: a structural and functional perspective. Front Microbiol
727 9:950.
728 53. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment
729 search tool. J Mol Biol 215:403-410.
730 54. Crow A, Greene NP, Kaplan E, Koronakis V. 2017. Structure and mechanotransmission
731 mechanism of the MacB ABC transporter superfamily. Proc Natl Acad Sci USA
732 114:12572-12577.
733 55. Sizikov S, Burgsdorf I, Handley KM, Lahyani M, Haber M, Steindler L. 2020.
734 Characterization of sponge-associated Verrucomicrobia: microcompartment-based sugar
735 utilization and enhanced toxin-antitoxin modules as features of host-associated
736 Opitutales. Environ Microbiol 22:4669-4688.
737 56. Erbilgin O, McDonald KL, Kerfeld CA. 2014. Characterization of a planctomycetal
738 organelle: a novel bacterial microcompartment for the aerobic degradation of plant
739 saccharides. Appl Environ Microbiol 80:2193-2205.
740 57. Lombard V, Ramulu HG, Drula E, Coutinho P, Henrissat B. 2014. The carbohydrate-
741 active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–D495.
742 58. Costa TRD, Felisberto-Rodrigues C, Meir A, Prevost MS, Redzej A, Trokter M,
743 Waksman G. 2015. Secretion systems in Gram-negative bacteria: structural and
744 mechanistic insights. Nat Rev Microbiol 13:343-359.

745 59. Visick KL, Stabb EV, Ruby EG. A lasting symbiosis: how Vibrio fischeri finds a squid
746 partner and persists within its natural host. Nat Rev Microbiol doi:10.1038/s41579-021-
747 00557-0.
748 60. Leonardi R, Jackowski S. 2007. Biosynthesis of pantothenic acid and coenzyme A.
749 EcoSal Plus 2:10.1128/ecosalplus.3.6.3.4.
750 61. Lackner G, Peters EE, Helfrich EJN, Piel J. 2017. Insights into the lifestyle of uncultured
751 bacterial natural product factories associated with marine sponges. Proc Natl Acad Sci
752 USA 114:E347-E356.
753 62. Chevrette MG, Gutierrez-Garcia K, Selem-Mojica N, Aguilar-Martinez C, Yanez-Olvera
754 A, Ramos-Aboites HE, Hoskisson PA, Barona-Gomez F. 2020. Evolutionary dynamics
755 of natural product biosynthesis in bacteria. Nat Prod Rep 37:566-599.
756 63. Chaisson MJ, Pevzner PA. 2008. Short read fragment assembly of bacterial genomes.
757 Genome Res 18:324-330.
758 64. Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Prjibelski
759 AD, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, Clingenpeel SR, Woyke T,
760 McLean JS, Lasken R, Tesler G, Alekseyev MA, Pevzner PA. 2013. Assembling single-
761 cell genomes and mini-metagenomes from chimeric MDA products. J Comput Biol
762 20:714-737.
763 65. Scholz M, Lo CC, Chain PSG. 2014. Improved assemblies using a source-agnostic
764 pipeline for metagenomic assembly by merging (MeGAMerge) of contigs. Sci Rep
765 4:6480.
766 66. Ruan J, Li L. 2019. Fast and accurate long-read assembly with wtdbg2. Nat Methods
767 17:155-158.
768 67. Vaser R, Sovic I, Nagarajan N, Sikic M. 2017. Fast and accurate de novo genome
769 assembly from long uncorrected reads. Genome Res 27:737-746.
770 68. Li P-E, Lo C-C, Anderson JJ, Davenport KW, Bishop-Lilly KA, Xu Y, Ahmed S, Feng
771 SH, Mokashi VP, Chain PSG. 2017. Enabling the democratization of the genomics
772 revolution with a fully integrated web-based bioinformatics platform. Nuc Acids Res
773 45:67-80.
774 69. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics
775 30:2068-2069.
776 70. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler
777 transform. Bioinformatics 25:1754-1760.
778 71. Blin K, Wolf T, Chevrette MG, Lu XW, Schwalen CJ, Kautsar SA, Duran HGS, Santos
779 E, Kim HU, Nave M, Dickschat JS, Mitchell DA, Shelest E, Breitling R, Takano E, Lee
780 SY, Weber T, Medema MH. 2017. antiSMASH 4.0-improvements in chemistry
781 prediction and gene cluster boundary identification. Nuc Acids Res 45:W36-W41.
782 72. Dong XL, Strous M. 2019. An integrated pipeline for annotation and visualization of
783 metagenomic contigs. Front Genetics 10:999.
784 73. Wu YW, Tang YH, Tringe SG, Simmons BA, Singer SW. 2014. MaxBin: an automated
785 binning method to recover individual genomes from metagenomes using an expectation-
786 maximization algorithm. Microbiome 2:26.
787 74. Wu YW, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm
788 to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605-607.

789 75. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the
790 quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome
791 Res 25:1043-1055.
792

793 **Table 1.** Metagenomic reads from 4 different samples were mapped back to the *Ca.* S.
794 palmerolidicus MAG.
795
796

	Synoicum adareanum samples			
	Bon-1c-2011	Del-2b-2011	Nor-2c-2007	Nor-2a-2007
Technology	PacBio CCS reads (1 cell)	PacBio CCS reads (3 cells)	454	Ion Torrent Proton
Number of reads	48,298	9,576	1,570,126	89,330,870
Mapped reads	21,591	3,522	23,993	15,979,084
Mapped reads (%)	44.70	36.78	1.53	17.89
Base coverage	99.98	99.89	90.15	99.98
Average fold	58.38	8.43	2.79	708.79
Gaps	2	3	1,734	8
Gap bases	644	4,618	422,870	774
SNPs	72	196	168	243
Indels	126	64	17	68

797
798

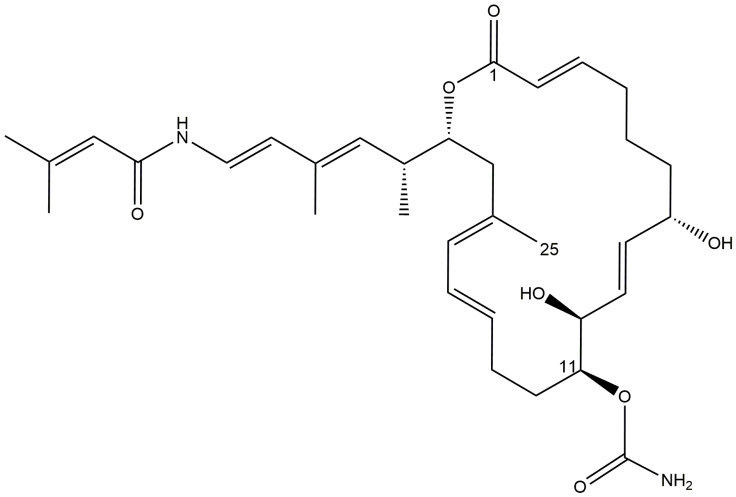
799 **Figure Legends**
800
801 **Figure 1.** Palmerolide A, a cytotoxic, macrolide with anti-melanoma activity is found in the
802 tissues of *Synoicum adareanum* in which a candidate biosynthetic gene cluster has been
803 identified. (A) *S. adareanum* occurs on rocky coastal seafloor habitats in the Antarctic; this study
804 focused on the region off-shore of Anvers Island in the Antarctic Peninsula. (B) Palmerolide A,
805 is the product of a hybrid PKS-NRPS system in which biosynthesis begins with a PKS starter
806 unit followed by incorporation of a glycine subunit by an NRPS module. Subsequent elongation,
807 cyclization and termination steps follow. Two additional features of the molecule include a
808 methyl group on C-25 and a carbamate group on C-11. (C) Five repeats encompassing candidate
809 palmerolide biosynthetic gene clusters were identified. The BGC (in blue) is defined as starting
810 with the NRP unit and ending at the carbamoyltransferase (green). Candidate palmerolide A
811 biosynthetic gene cluster BGC4 was identified from initial metagenome library assemblies. The
812 other four clusters were identified following a third round of sequencing, assembly and manual
813 finishing. Primary BGC coding sequences (CDS) and a conserved tailoring cassette are in blue.
814 Light blue CDS are an ATP transporter with homology to an antibiotic transporter, SyrD. All
815 black CDS are repeated amongst the BGCs. Orange CDS are transposase/integrase domains.
816 Gray CDS are unique, non-repeated; and in BGC2 and 5, the unique CDS encode transposases,
817 distinct from the predicted amino acid sequences of those in orange. The red lines associated
818 with Contig 9 indicate targeted quantitative PCR regions.
819
820 **Figure 2.** Abundances of real time PCR-targeted coding regions in the candidate pal biosynthetic
821 gene cluster in Antarctic ascidian samples. (A) Gene copies estimated for three targeted coding
822 regions (Acyltransferase, AT1; 3-hydroxy-methyl-glutaryl coenzyme A synthase, HCS; and the
823 condensation domain of a non-ribosomal peptide synthase; NRPS) in the candidate *pal*
824 biosynthetic gene cluster surveyed over 63 DNA extracts derived from microbial cell
825 preparations enriched from the Antarctic ascidian *Synoicum adarenum.* Nine samples were
826 collected at each of seven sites: Bon, Bonaparte Point; Del, Delaca Island; Jan, Janus Island, Kil,
827 Killer Whale Rocks; Lag, Laggard Island, Lit, Litchfield Island; Nor, Norsel Point (27). (B)
828 Relationship between gene copy number for the three gene targets and the 16S rRNA gene ASV
829 occurrences of Opitutaceae-related ASV_15 across a 63 *S. adareanum* microbial DNA sample
830 set. * indicates samples Bon-1C-2011 and Del-2b-2011 that were selected for PacBio
831 sequencing.
832
833 **Figure 3**. Genome maps of assembled MAG, *Candidatus* Synoicohabitans palmerolidicus and
834 evidence of multi-copy biosynthetic gene clusters. (a) The 4,297,084 bp gene map is oriented to
835 dnaA at the origin. One possible assembly scenario of the *Ca.* S. palmerolidicus genome is
836 shown as the order of the contigs and palmerolide BGC's are not currently known. In addition to
837 the 5 BGCs, three other internally repetitive regions were identified (15.3, 17.0 and 27.4 Kbp).
838 The genes and orientation are shown with blue and tRNA indicated in red. (b) To demonstrate
839 depth of coverage outside and inside the BGC regions, CCS reads from sample Bon-1C-2011
840 and Del-2b-2011 were mapped to a 167.6 Kbp region. The profile extends 40 Kbp into the
841 genome on either side of the BGC where depth of coverage averages 60 fold, while in the BGC
842 depth of coverage varies across the BGC given differences in cover across the BGC, the highest
843 cover is 5X, or ~ 300 fold, supporting the finding of 5 repeats encoding the BGC.
844

845 **Figure. 4.** Maximum likelihood phylogenomic tree showing 48 verrucomicrobia genomes.
846 Phylogenomic relationship of *Candidatus* Synoicohabitans palmerolidicus (*Opitutaceae* bin 8)
847 with respect to other mostly marine, and host-associated verrrucomicrobia subdivision 4 and
848 other genomes. The tree is based on 16 concatenated ribosomal proteins (5325 amino acids)
849 common across 48 Verrucomicrobia genomes. Distance was estimated with RAxML with 300
850 bootstrap replicates. Symbols designate environmental origins of the organisms: free-living are
851 represented by circles: light blue - marine, green– freshwater, red – hydrothermal mud, brown –
852 soils. Host-associated taxa from marine systems with blue diamonds, and from terrestrial systems
853 with black diamonds.
854
855 **Supplemental Material**
856
857 **Supplemental Materials and Methods**
858 Details concerning ascidian sample collections, sample processing and high molecular weight
859 DNA extraction, metagenome sequencing, metagenome binning, bin taxonomic and functional
860 classification, Real Time PCR, manual assembly procedures, annotation and phylogenomic
861 analyses and associated references are provided in the supplemental materials and methods text
862 section.
863
864 **Supplemental Figure Legends**
865
866 **Figure S1**. Ribosomal RNA-based phylogenetic tree. Sequences were selected from
867 Verrucomicrobia (Phylum) and mostly Opitutaceae (family) genomes represented by isolates,
868 metagenome assembled genomes and single cell assembled genomes from host associated and
869 marine ecosystems. RAXML tree is based on 1,636 bases; 250 bootstraps were run, and values >
870 49 are shown at nodes. *Ca*. S. palmerolidicus (*Opitutaceae* bin 8) is indicated with an orange
871 diamond symbol. Symbols designate environmental origins of the organisms: free-living are
872 represented by circles: light blue - marine, green– freshwater, red – hydrothermal mud, brown –
873 soils. Host-associated taxa from marine systems with blue diamonds, and from terrestrial systems
874 with black diamonds.
875
876 **Figure S2**. Phylogenetic relationships of homologs of (a) MacB CDS and (b) celluase
877 glycosylhydrolase family 5 enzymes identified in the Ca. S. palmerolidicus genome and closest
878 neighboring sequences that were sourced from GenBank using BLAST. Maximum likelihood
879 analysis was conducted using RAxML v. 8.2.12 using the PROTGAMMALG model with 550
880 bootstrap replicate trees calculated for (a) and 1000 for (b). Bootstrap values > 49 are shown
881 which represent percent of times the topology was found.
882
883 **Figure S3**. Bacterial microcompartment (BMC) super loci identified in the *Ca.* S. palmerolidicus
884 MAG. Conserved structural elements identified with Pfam and Swiss-Prot annotations were
885 consistent with other PV-BMC loci which are associated carbohydrate utilization domains.
886 Transcriptional regulatory genes are indicated in green, conserved PV-BMC enzymes in blue, the
887 BMC proteins in red, enzymes associated with rhamnose metabolism in violet and those with
888 lactate utilization in purple.
889
890 **Supplemental Table Legends**

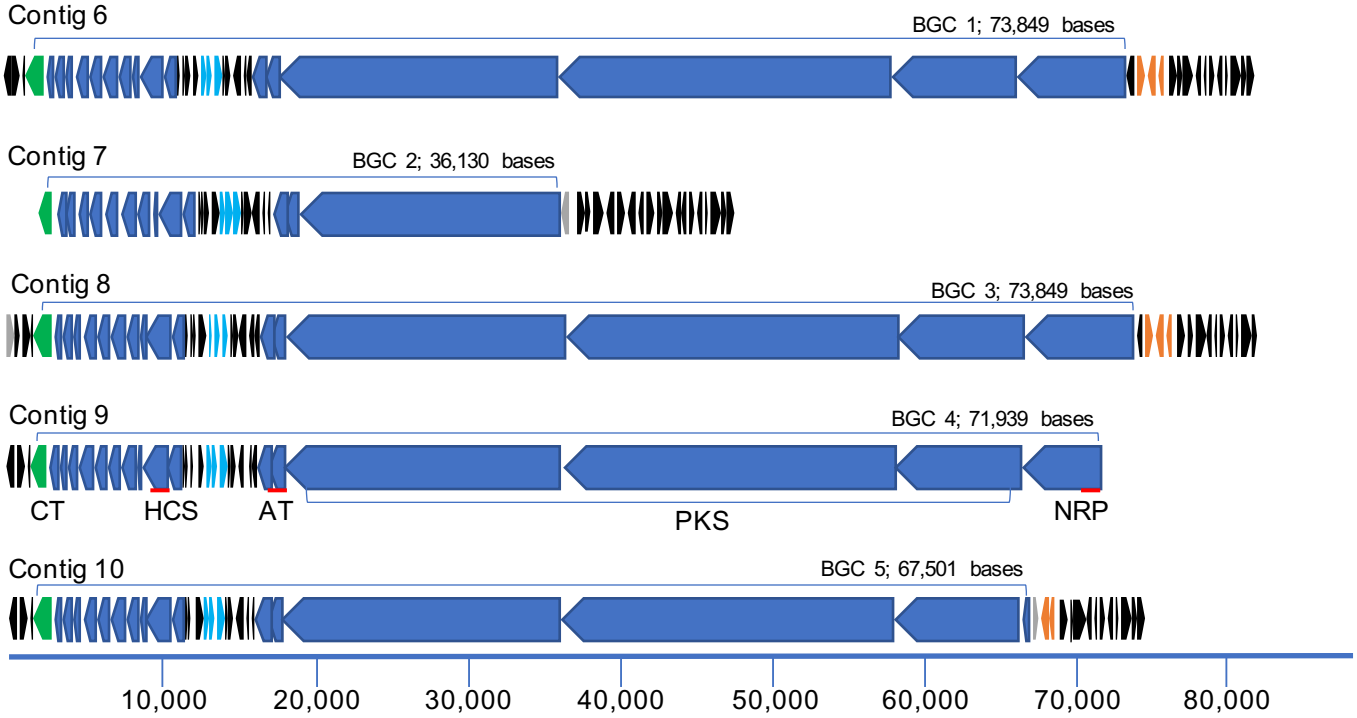
891
892 **Table S1.** Metagenome sequencing metadata for *S. adareanum* microbiome. Host lobe identities
893 indicate sample site (Nor, Norsel, Bon, Bonaparte Point, Del, DeLaca Island; followed by lobe
894 designation, and year sampled). Assembly 1 joined two data sets using mega-merged contigs
895 from 454 and Ion Proton sequence data sets. Assembly 2 included Assembly 1 and PacBio
896 assemblies in addition to the manually assembled *Ca.* S. palmerolidicus MAG.
897
898 **Table S2**. Assembly 1 antiSMASH results for the 102 contigs > 40 Kbp. The 72.1 and 78.1
899 regions encoded biosynthetic gene clusters presented two putative clusters that were further
900 investigated using QPCR to target samples with high BGC copy numbers for a subsequent round
901 of metagenome sequencing.
902
903 **Table S3**. Quality checks and taxonomic classification using CheckM and GTDB-Tk for
904 Opitutaceae-associated bins at different stages of analysis (most recent, highest quality on the
905 left-most column and initial Coassembly 1 and bin 4 in the right-most column). The *Ca*. S.
906 palmerolidicus-2 column reflects identifications (MetaERG annotation) of many of the missing
907 markers in *Ca*. S palmerolidicus-1. The Opitutaceae bin 8 – clean CoAssembly 2 data set reflects
908 an effort to remove non-Opitutaceae sequences from the bin (see supplemental material,
909 methods). CoAssembly 1 bins 1 and 2 included the palmerolide A biosynthetic gene cluster and
910 were otherwise populated by short contigs, and were taxonomically unidentified. CoAssembly 1
911 bin 4 included Opitutales-related contigs including a SSU rRNA gene sequence.
912
913 **Table S4**. GC percent for Opitutaceae-family representatives. Different genera are represented
914 with different colors. Marine Optitutaceae metagenome assembled genomes are indicated with *.
915 Source and species nomenclature: GTDB release 05-RS95. The GC content for *Ca.*
916 Synoicihabitans palmerolidicus is 58.7%.
917
918 **Table S5.** Quantitative PCR gene targets and primers. Primers targeting three gene targets were
919 used in this study Primer IDs are based on bases from the beginning of the CDS. A GBlocks
920 synthetic DNA positive control was used for estimating copy number[†].
921
922 **Table S6**. Conserved markers used in phylogenomic analysis. Integrated microbial genomes
923 (IMG) and/or GenBank accession numbers are listed. Shaded markers were those used for
924 phylogenomic analysis (Fig. 4).
925

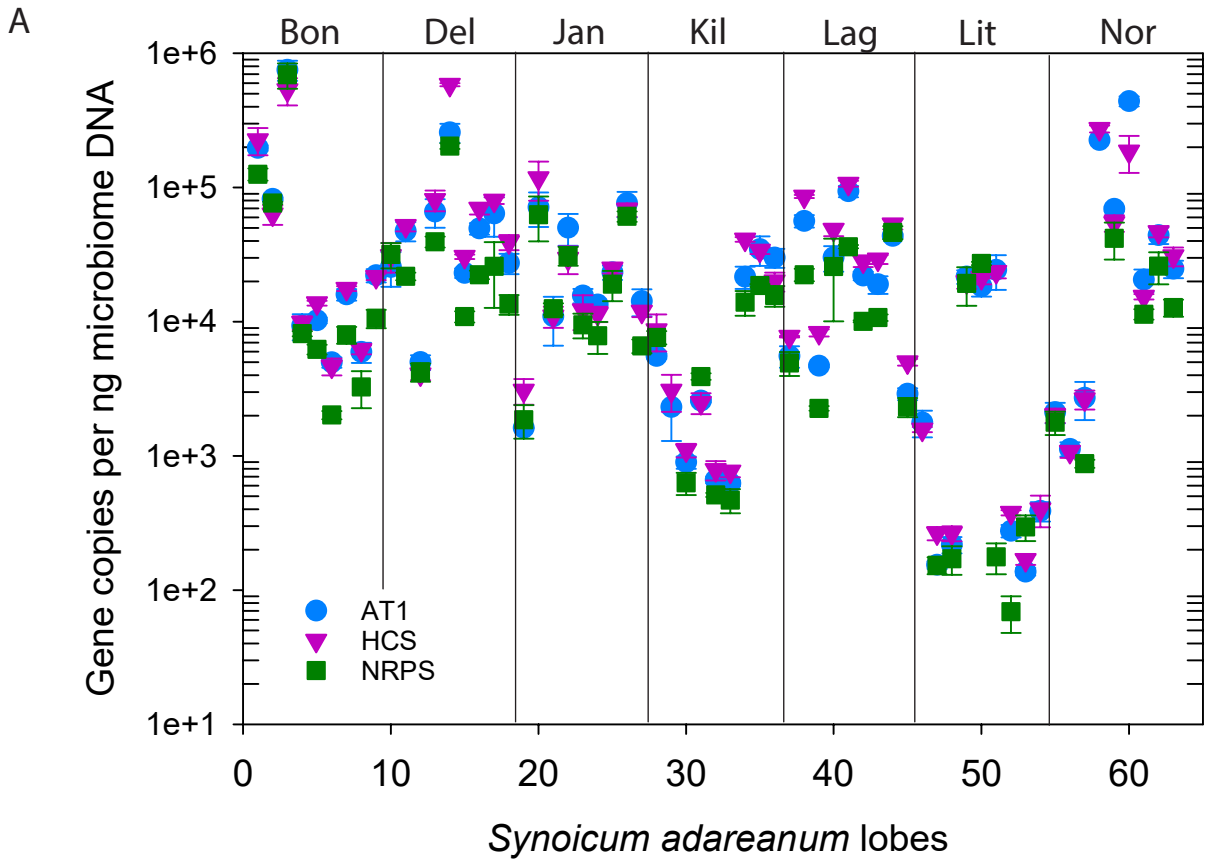
A

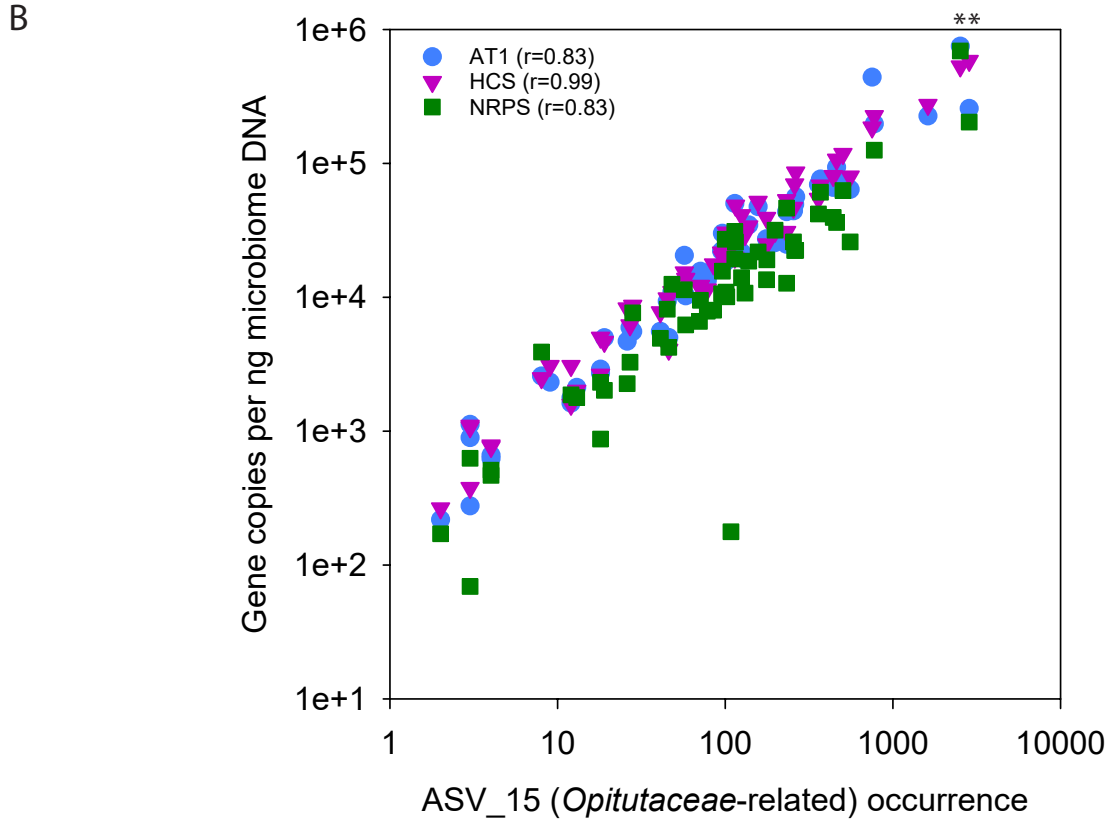
B


O
1
O
H
N
O
25
OH
HO
11
O
O
NH2

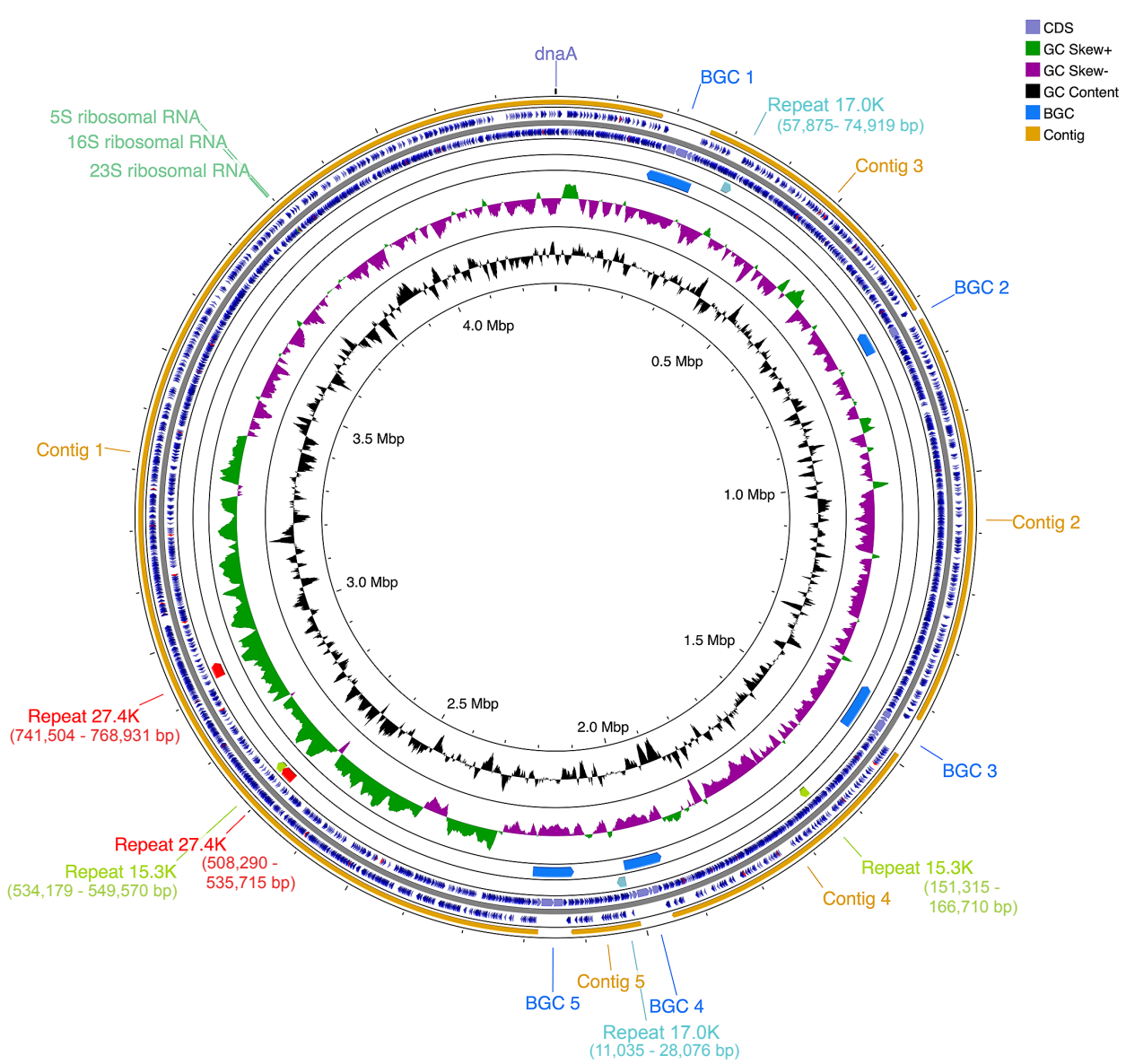
C


Contig 6
BGC 1; 73,849 bases
Contig 7
BGC 2; 36,130 bases
Contig 8
BGC 3; 73,849 bases
Contig 9
BGC 4; 71,939 bases
CT
HCS
AT
PKS
NRP
Contig 10
BGC 5; 67,501 bases
10,000
20,000
30,000
40,000
50,000
60,000
70,000
80,000

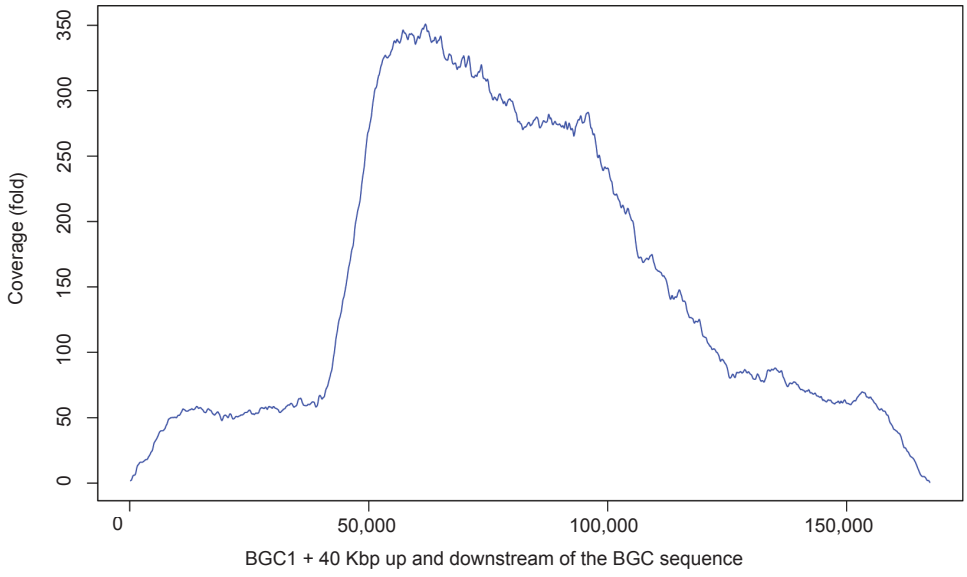

A
Bon
Del
Jan
Kil
Lag
Lit
Nor
Gene copies per ng microbiome DNA
1e+6
1e+5
1e+4
1e+3
1e+2
1e+1
AT1
HCS
NRPS
0
10
20
30
40
50
60
Synoicum adareanum lobes

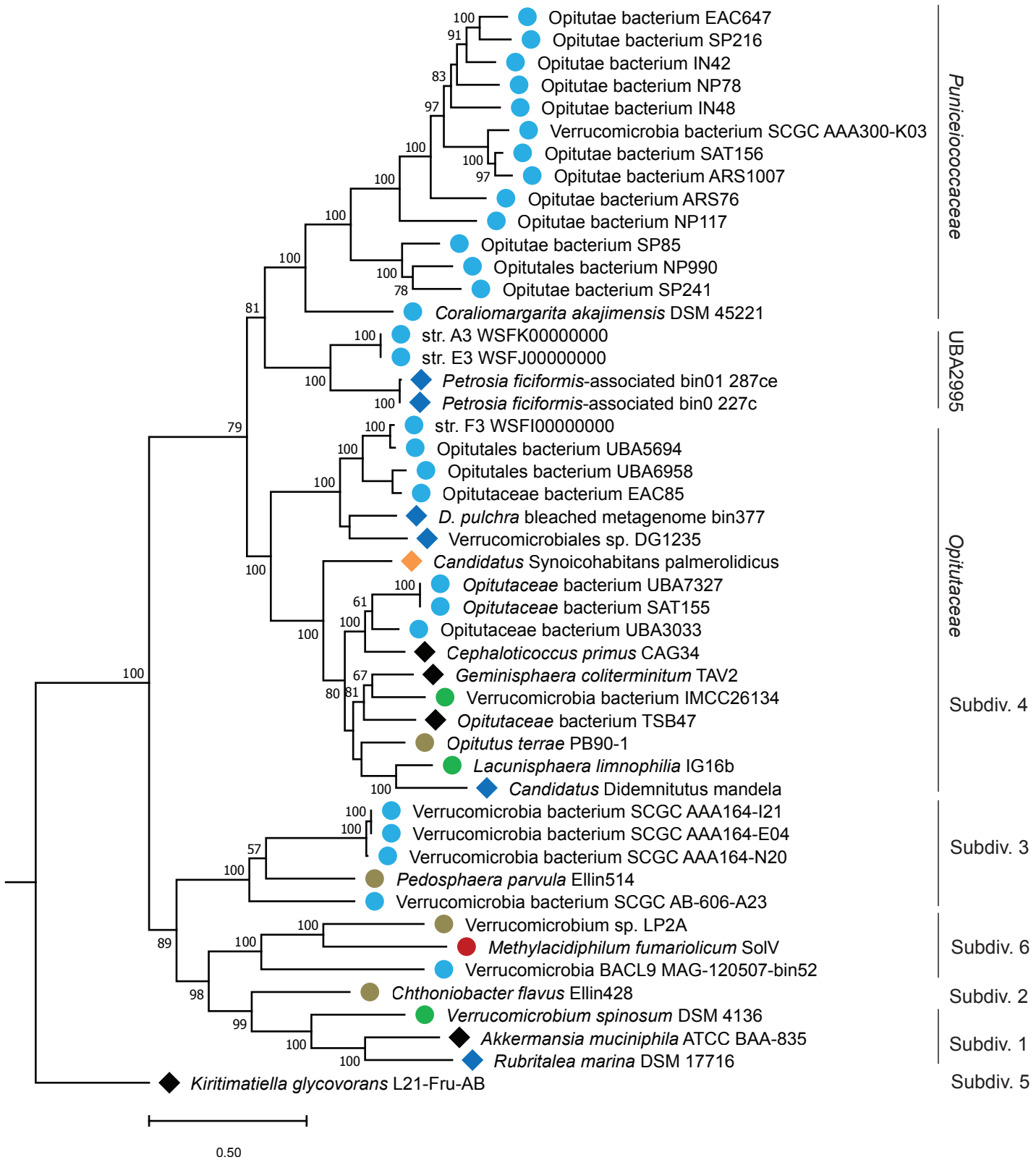

B
**
AT1 (r=0.83)
HCS (r=0.99)
NRPS (r=0.83)
Gene copies per ng microbiome DNA
1e+6
1e+5
1e+4
1e+3
1e+2
1e+1
1
10
100
1000
10000
ASV_15 (Opitutaceae-related) occurrence

A


CDS
GC Skew+
GC Skew-
GC Content
BGC
Contig
dnaA
BGC 1
Repeat 17.0K
(57,875- 74,919 bp)
Contig 3
BGC 2
Contig 2
BGC 3
Repeat 15.3K
(151,315 - 166,710 bp)
Contig 4
Repeat 17.0K
(11,035 - 28,076 bp)
BGC 4
Contig 5
BGC 5
Repeat 27.4K
(508,290 - 535,715 bp)
Repeat 15.3K
(534,179 - 549,570 bp)
Repeat 27.4K
(741,504 - 768,931 bp)
Contig 1
23S ribosomal RNA
16S ribosomal RNA
5S ribosomal RNA
0.5 Mbp
1.0 Mbp
1.5 Mbp
2.0 Mbp
2.5 Mbp
3.0 Mbp
3.5 Mbp
4.0 Mbp

B


Coverage (fold)
0
50
100
150
200
250
300
350
0
50,000
100,000
150,000
BGC1 + 40 Kbp up and downstream of the BGC sequence


Opitutae bacterium EAC647
Opitutae bacterium SP216
Opitutae bacterium IN42
Opitutae bacterium NP78
Opitutae bacterium IN48
Verrucomicrobia bacterium SCGC AAA300-K03
Opitutae bacterium SAT156
Opitutae bacterium ARS1007
Opitutae bacterium ARS76
Opitutae bacterium NP117
Opitutae bacterium SP85
Opitutales bacterium NP990
Opitutae bacterium SP241
Coraliomargarita akajimensis DSM 45221
str. A3 WSFK00000000
str. E3 WSFJ00000000
Petrosia ficiformis-associated bin01 287ce
Petrosia ficiformis-associated bin0 227c
str. F3 WSFI00000000
Opitutales bacterium UBA5694
Opitutales bacterium UBA6958
Opitutaceae bacterium EAC85
D. pulchra bleached metagenome bin377
Verrucomicrobiales sp. DG1235
Candidatus Synoicohabitans palmerolidicus
Opitutaceae bacterium UBA7327
Opitutaceae bacterium SAT155
Opitutaceae bacterium UBA3033
Cephaloticoccus primus CAG34
Geminisphaera coliterminitum TAV2
Verrucomicrobia bacterium IMCC26134
Opitutaceae bacterium TSB47
Opitutus terrae PB90-1
Lacunisphaera limnophilia IG16b
Candidatus Didemnitutus mandela
Verrucomicrobia bacterium SCGC AAA164-I21
Verrucomicrobia bacterium SCGC AAA164-E04
Verrucomicrobia bacterium SCGC AAA164-N20
Pedosphaera parvula Ellin514
Verrucomicrobia bacterium SCGC AB-606-A23
Verrucomicrobium sp. LP2A
Methylacidiphilum fumariolicum SolV
Verrucomicrobia BACL9 MAG-120507-bin52
Chthoniobacter flavus Ellin428
Verrucomicrobium spinosum DSM 4136
Akkermansia muciniphila ATCC BAA-835
Rubritalea marina DSM 17716
Kiritimatiella glycovorans L21-Fru-AB
Puniceiococcaceae
UBA2995
Opitutaceae
Subdiv. 4
Subdiv. 3
Subdiv. 6
Subdiv. 2
Subdiv. 1
Subdiv. 5
0.50