Main Manuscript for

Palmerolide PKS-NRPS gene clusters characterized from the microbiome of an Antarctic ascidian

Nicole E. Avalon;[1] Alison E. Murray;[2] Hajnalka E. Daligault;[3] Chien-Chi Lo;[3] Karen W. Davenport;[3] Armand E.K. Dichosa;[3] Patrick S.G. Chain;[3] Bill J. Baker[1]

[1]Department of Chemistry, University of South Florida, Tampa, FL 33620
[2]Division of Earth and Ecosystem Sciences, Desert Research Institute, Reno, NV 89512
[3]Los Alamos National Laboratory, Los Alamos, NM 87545

Corresponding author:

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

Author Contributions: This work was the result of a team effort in which the following contributions are recognized: conceptualization, A.E.M., P.S.G.C., and B.J.B.; methodology and experimentation, N.E.A., A.E.M., H.E.D., C.L., P.S.G.C., B.J.B; data analysis, N.E.A., A.E.M., H.E.D, C.L., K.W.D., A.E.K.D., P.S.G.C, B.J.B; data curation, N.E.A., A.E.M., H.E.D, C.L., P.S.G.C, B.J.B; original draft preparation, N.E.A., B.J.B., A.E.M., H.E.D.; review and editing, N.E.A., B.J.B., A.E.M., C.L., H.E.D., P.S.G.C., A.E.K.D; funding acquisition, A.E.M., P.S.G.C., and B.J.B. All authors have read and agreed to the published version of the manuscript.

Competing Interest Statement: The authors have no competing interests to disclose.

Classification: Biological sciences, environmental sciences and ecology

Keywords: microbial biosynthesis, secondary metabolites, Antarctic microbiology, *trans*-AT Type I polyketide synthase, marine natural products

This PDF file includes:

Main Text.
Figures 1 to 4.

Abstract

Polyketides are a complex family of natural products that often serve competitive or pro-survival purposes but can also demonstrate bioactivity in human diseases as, for example cholesterol lowering agents, anti-infectives, or anti-tumor agents. Marine invertebrates and microbes are a rich source of polyketides. Palmerolide A, a polyketide isolated from the Antarctic ascidian *Synoicum adareanum,* is a vacuolar-ATPase inhibitor with potent bioactivity against melanoma cell lines. The biosynthetic gene clusters (BGC) responsible for production of secondary metabolites are encoded in the genomes of the producers as discrete genomic elements. A putative palmerolide BGC was identified from a *S. adareanum* metagenome based on a high degree of congruence with a chemical structure-based retrobiosynthetic prediction. Protein family homology analysis, conserved domain searches, and active site and motif identification were used to confirm the identity and propose the function of the 75 kb *trans*-acyltransferase (AT) polyketide synthase-non-ribosomal synthase (PKS-NRPS) domains responsible for the synthesis of palmerolide A. Though PKS systems often act in a predictable co-linear sequence, this BGC includes multiple *trans*-acting enzymatic domains, a non-canonical condensation termination domain, a bacterial luciferase-like monooxygenase (LLM), and multiple copies found within the metagenome-assembled genome (MAG) of *Candidatus* Synoicohabitans palmerolidicus. Detailed inspection of the five highly similar pal BGC copies suggests the potential for biosynthesis of other members of the palmerolide chemical family. This is the first delineation of a biosynthetic gene cluster from an Antarctic species. These findings have relevance for fundamental knowledge of PKS combinatorial biosynthesis and could enhance drug development efforts of palmerolide A through heterologous gene expression.

Significance Statement

Complex interactions exist between microbiomes and their hosts. Increasingly, defensive metabolites that have been attributed to host biosynthetic capability are now being recognized as products of associated microbes. These unique metabolites often have bioactivity in targets of human disease and can be purposed as pharmaceuticals. The molecular machinery for production of palmerolide A, a macrolide that is potent and selective against melanoma, was discovered as a genomic cluster in the microbiome of an Antarctic ascidian. Multiple non-identical copies of this genomic information provide clues to differences in specific enzymatic domains and point to Nature's ability to perform combinatorial biosynthesis *in situ*. Harnessing this genetic information may pave a path for development of a palmerolide-based drug.

Introduction

Marine invertebrates such as corals, sponges, mollusks, and ascidians, are known to be a rich source of bioactive compounds (1). Due to their sessile or sluggish nature, chemical defenses such as secondary metabolites are often key to their survival. Many compound classes are represented among benthic invertebrates including terpenes, nonribosomal peptide synthase (NRPS) products, ribosomally synthesized and post-translationally modified peptides (RiPPs), and polyketides. It is estimated that that over 11,000 secondary metabolites from marine and terrestrial environments understood to be of polyketide synthase (PKS) and NRPS origin have been isolated and described (2). Biosynthetic gene clusters (BGCs) exist as a series of genomic elements that encode for megaenzymes responsible for production of these secondary metabolites. BGCs can have distinct nucleotide composition properties such as codon usage and guanine-cytosine content that do not match the remainder of the genome (3, 4), suggesting a mechanism of transfer from organisms that are only distantly related, perhaps even from interkingdom horizontal gene transfer (5, 6). Interestingly, the BGCs for many natural products isolated from marine invertebrates are found in the host-associated microbiota, reflecting the role of these compounds in symbiosis (7).

Polyketides are a complex family of natural products produced by multi-modular PKS enzymes that are related to, but evolutionarily divergent from, fatty acid synthases (8). They often

possess long carbon chains with varied degrees of oxidation, can contain aromatic components, and may be either cyclic or linear. It is estimated that of the polyketides that have been isolated and characterized, 1% have potential biological activity against human diseases, making this class of compounds particularly appealing from a drug discovery and development standpoint (9). This potential for use as pharmaceuticals is approximately five times greater than for compounds of all other natural product classes (9). Many polyketides are classified as macrolides, which are large-ring lactones that are pharmaceutically relevant due to a number of biological actions, including, for example, targeting the cytoskeleton, ribosomal protein biosynthesis, and vacuolar type V-ATPases (10–13). V-ATPases are responsible for acidification of cells and organelles via proton transport across membranes, including those of lysosomes, vacuoles, and endosomes. These enzymes appear to have an impact on angiogenesis, apoptosis, cell proliferation, and tumor metatastisis (10). A number of marine macrolides inhibit V-ATPases, including lobatamides, chondropsins, iejimalides, and several of the palmerolides (14–18).

There are three types of PKS systems: Type I, which are comprised of non-iteratively acting modular enzymes that lead to progressive elongation of a polyketide chain, Type II, which contain iteratively acting enzymes that biosynthesize polycyclic aromatic polyketides, and Type III which possess iteratively-acting homodimeric enzymes that often result in monocyclic or bicyclic aromatic polyketides (19). Type I PKS systems can be subdivided into two groups, depending upon whether the acyl transferase (AT) modules are encoded co-linearly, as in *cis*-AT Type I PKS, or physically distinct from the megaenzyme, as in *trans*-AT Type I PKS. The concept of co-linearity refers to the parallel relationship between the order of enzymatic modules and the functional groups present in the growing polyketide chain (20). In *trans*-AT systems, AT domains may be incorporated in a mosaic fashion through horizontal gene transfer (20). This introduces greater molecular architectural diversity over evolutionary time, as one clade of *trans*-ATs may select for a malonyl-CoA derivative, while the *trans*-AT domains in another clade may select for unusual or functionalized subunits (21). Additionally, recombination, gene duplication, and conversion events can lead to further diversification of the resultant biosynthetic machinery (21). Predictions regarding the intrinsic relationship between a secondary metabolite of interest, the biosynthetic megaenzyme, and the biosynthetic gene cluster (BGC) can be harnessed for natural product discovery and development (22–24).

In the search for new and bioactive chemotypes as inspiration for the next generation of drugs, underexplored ecosystems hold promise as biological and chemical hotspots (25). The vast Southern Ocean comprises one-tenth of the total area of Earth's oceans and is largely unstudied for its chemodiversity. The coastal marine environment of Antarctica experiences seasonal extremes in, for example, ice cover, light field, and food resources. Taken with the barrier to migration imposed by the Antarctic Circumpolar Current and the effects of repeated glaciation events on speciation, a rich and endemic biodiversity has evolved, with consequent potential for new chemodiversity (25–27).

Palmerolide A (**Fig. 1**) is the principal secondary metabolite isolated from *Synoicum adareanum*, an ascidian which can be found in abundance at depths of 10 to 40 m in the coastal waters near Palmer Station, Antarctica (17). Palmerolide A is a macrolide polyketide that possesses potent bioactivity against malignant melanoma cell lines while demonstrating minimal cytotoxicity against other cell lines (17). The NCI's COMPARE algorithm was used to correlate experimental findings with a database for prediction of the biochemical mechanism by identifying the mechanism of action of palmerolide A as a vacuolar-ATPase (V-ATPase) inhibitor (28). Downstream effects of V-ATPase inhibition include an induction of both hypoxia induction factor-1α and autophagy (17, 29). Increased expression of V-ATPase on the surface of metastatic melanoma cells perhaps explain palmerolide A's selectivity for UACC-62 cell lines over all cell types (30). Despite the relatively high concentrations of palmerolide A in the host tissue (0.49–4.06 mg palmerolide A x g−1 host dry weight) (31), isolation of palmerolide A from its Antarctic source in mass sufficient for drug development it is not ecologically nor logistically feasible. Synthetic strategies for the synthesis of palmerolide A have been reported (6–13), though a clear pathway to quantities needed for drug development has been elusive. Therefore, there is substantial interest in identifying the biosynthetic gene cluster (BGC) responsible for palmerolide A production.

Our approach to identify the palmerolide BGC (*pal* BGC) began with the characterization of the host-associated microbiome (39). The core microbiome for *Synoicum adareanum*, a persistent cohort of bacteria present in many individual ascidians, was established through analysis of occurrence of distinct amplicon sequence variants (ASV) from the iTag sequencing of the Variable 3-4 regions of the bacterial 16S rRNA (31). This ultimately led to the evaluation of the metagenome and then the assembly of the metagenome assembled genome (MAG) of *Candidatus* Synoicohabitans palmerolidicus, a verrucomicrobium in the family *Opitutaceae* (40). Contained within the genome are five non-identical copies of the *pal* BGC, of which, three correlate with palmerolides, the structures of which have been previously discovered in this macrolide family (**Fig. 1**). Here we report a detailed analysis of the *pal* BGCs.

Results and Discussion

A. Retrobiosynthetic Scheme for Palmerolide A

A retrobiosynthetic scheme of the *pal* BGC was developed based on the chemical structure for palmerolide A, including modules consistent with a hybrid PKS-NRPS with tailoring enzymes for key functional groups (**Fig. 1**). We hypothesized that the initial module would be PKS-like in nature to utilize 3-methylcrotonic acid as the starter unit followed by a NRPS domain for the incorporation of glycine. PKS elongation was predicted to be an 11-step sequence resulting in 22 contiguous carbons. Modifying enzymes that are encoded co-linearly were predicted to create the architectural diversity with olefin placement, reduction of certain carbonyl groups to secondary alcohols, and full reduction of other subunits. In addition, carbon methyltransferases (cMTs) were predicted to catalyze the placement of methyl groups C-26 and C-27 from *S*-adenosylmethionine (SAM).

Several key structural features proposed to result from the action of *trans*-acting enzymes are present. As seen in the kalimantacins (41), the carbamate on C-11 was hypothesized to be installed by a carbamoyl transferase (CT). The C-25 methyl group is located on C-17, a carbon atom in α-position of the acetate subunit. This structural feature suggests its origin from hydroxymethylglutamoyl-CoA Synthetase (HCS) catalysis, rather than the more common SAM-mediated methylation at the acetate β-position, which appears to be the origin of the C-26 and C-27 methyl groups. Precedence for HCS-mediated β-branch formation is found in the biosynthetic pathways for of the jamaicamides, bryostatins, curacin A, oocydin, pederin, and psymberin, among others (42–46). The hydroxy group on C-10, in the α-position of the acetate subunit, was hypothesized to result from incorporation of hydroxymalonic acid or the action of a hydroxylase on a malonyl-CoA monomer.

B. Proposed Biosynthesis and Architecture of the Putative *pal* Biosynthetic Gene Cluster

Synoicum adareanum collected near Palmer Station, Antarctica was processed to separate eukaryotic from prokaryotic cells of the host microbiome. Iterations of metagenomic analysis (40) yielded a large number of BGCs when analyzed via the antiSMASH annotation platform. A candidate cluster with multiple copy numbers was identified and has excellent congruence with retrobiosynthetic predictions, including the key genomic markers outlined above that are likely involved in biosynthesis of palmerolide A. Integration of BGC annotations with information from protein family homology analysis, conserved domain searches, active site and motif identification, supported the hypothesis that this putative 75 kb BGC is responsible for palmerolide A production. The core scaffold was explained by the NRPS and *trans*-AT PKS hybrid system. In addition, each of the tailoring enzymes that are expected for the distinct chemical features are encoded in the genome. The BGC for palmerolide A is comprised of 14 core biosynthetic modules and 25 genes in a single operon of 74,655 bases (**Fig. 2**). Fourteen modules are co-linear and two *trans*-AT domains (modules 15 & 16) follow the core biosynthetic genes. Additional *trans*-acting genes

contribute to backbone modifications with at least one gene contributing to post-translational tailoring (**Fig. 3**).

An unusual starter unit and NRPS domains of *palA.* The initial core biosynthetic domains of *palA* (modules 1 and 2) encode for the requisite acyl carrier proteins (ACP) (**Fig. 2**). Encoded in module 1, there are three ACPs in tandem, which could serve to promote an increase in metabolite production (47). The second in series is an ACP-β containing the conserved domain sequence GXDS (48) and is likely the acceptor of a starter unit containing a β-branch, consistent with the proposed starter unit for palmerolide A, 3-methylcrotonic acid. While both *trans*-acting ATs, PalE and PalF, (**Fig. 2**) possess the catalytic active site serine which is key for the proper positioning of subunit within the hydrophobic cleft of the active site (8, 49), only the first AT, PalE, has a characteristic motif that includes an active site phenylalanine, conferring specificity for malonate selection (50). The AT selecting the methylcrotonic acid starter unit is likely the second of the two *trans*-AT domains (PalF), which lacks definitive specificity for malonyl-CoA. In support, *trans*-ATs have demonstrated affinity for a wider range of substrates than their *cis*-acting counterparts (21, 51). 3-Methylcrotonyl-CoA is an intermediate of branched-chain amino acid catabolism in leucine degradation; intermediates of which can be diverted to secondary metabolite production (52). Interestingly, an upregulation of leucine catabolic pathways can be seen as a mechanism of cold adaption to maintain cell wall fluidity via production of branched chain fatty acids, a factor which may be at play in this Antarctic microorganism (52, 53). Condensation (C) and adenylation (A) domains accompany the ACP in module 2. Signature sequence information and NRPSPredictor2 method (54, 55) of the A domain are consistent with selection of a glycine residue. These domains incorporate the amino acid residue, resulting in the addition of a nitrogen and two carbons in this step of the biosynthesis of palmerolide A.

Contiguous PKS chain and *trans*-acting enzymes at site of action for *palB – palD*. The contiguous carbon backbone of palmerolide A arises from 11 cycles of elongation from a series of modules with a variety of enzymatic domains that include an ACP, KS, and associated genes that establish the oxidation state of each subunit (**Fig 2**). The first module of *palB* (module 3) includes a dehydratase (DH) and cMT domains, a sequence which results in a chain extension modification to an α,β-unsaturated thioester, a result of the action of the encoded DH. The subsequent cMT methylation is consistent with an *S*-adenosylmethionine (SAM)-derived methyl group, as expected for C-27 in palmerolide A. Module 4, spanning the end of *palB* and beginning of *palC*, includes a DH, a ketoreductase (KR), an epimerization (E), and another cMT domain. This cluster of domains is predicted to result in the methyl-substituted conjugated diene of the macrolide tail (C-19 through C-24, C-26 on palmerolide A). The epimerization domain is hypothesized to set the stereochemistry at C-20.

The substrate critical for macrolactonization of the polyketide is the C-19 hydroxy group, a result of the KS, E, and KR domains encoded in module 5 (**Fig. 2**). Additionally, a domain initially annotated as a dehydratase (DHt) at this location may contribute to the final cyclization and release of the molecule from the megaenzyme by assisting the terminal C domain with ring closure (48). This sequence shows some homology with condensation domains and does not possess the hotdog fold that is indicative of canonical dehydratases, and therefore, may not truly represent a DH (56). Alternatively, this domain could be responsible for the olefin shift to the β,γ-position, as seen in bacillaene and ambruticin biosynthesis (57, 58). Finally, the E domain is hypothesized to establish the stereochemistry at C-19.

Only an enoyl-CoA hydratase (ECH) in addition to standard ACP and KS is encoded in module 6 which would lead to a ketone function; based on our retrobiosynthetic analysis, the resulting ketone at C-17 is the necessary substrate for HCS-catalyzed β-branch formation, resulting in the C-25 methyl group on C-17. The HCS cassette (PalK through PalO) is comprised of a series of *trans*-acting domains, including an ACP, an HCS, a free KS, and 2 additional ECH modules (**Fig. 2**). The HCS cassette can act while the elongating chain is tethered to an ACP module, rather than after cyclization and release (59, 60). The two ECHs in the HCS cassette along with the ECH encoded in-line with the core biosynthetic genes would be responsible for isomerization of a

248 terminal methylene to the observed internal olefin. An HCS cassette formed by the combination of
249 a *trans*-KS and at least one ECH module with an HCS domain is reported in several other bacterial
250 BGCs such as bryostatin 1, calyculin A, jamaicamide, mandelalide, phormidolide, and psymberin
251 (43, 46, 48, 61–63). The domain structure for the HCS cassettes has a remarkably high degree of
252 syntony (64).
253 There is significant similarity in the domain structure of module 7, module 10, and module 13,
254 which each include a KS, DH, and KR (**Fig. 2**). The olefin that arises from the action of module 7,
255 concomitant with carbon chain elongation, is conjugated with the C16-C17 olefin adjacent the C-
256 17 β-branch. Modules 10 and 13 have similar enzymatic composition to 7 and are likely responsible
257 for Δ^8 and Δ^2 olefins. The combination of KR and DH domains are also found in modules 8 and 12;
258 however, in concert with an unidentified *trans*-acting ER domain, these olefins would be reduced
259 to fully saturated monomeric subunits. There are some examples of *trans*-acting ER domains
260 carrying out this function, including OocU in oocydin, SorN in sorangicin, and MndM in mandelalide
261 (45, 63, 65), while in other systems, such as corallopyronin and leinamycin, the reductions of the
262 olefins are largely unexplained (66, 67). The reduction by a *trans*-acting enzyme often occurs while
263 the elongating polyketide is tethered to the megaenzyme, as evidenced by the downstream
264 specificity of the KS module for Claisen-type condensation with subunits containing single or double
265 bonds (65).
266 The genetic architecture for the biosynthesis of two functional groups, essential for bioactivity,
267 is encoded in module 9 (**Fig. 2**); structure-activity relationship studies demonstrate the importance
268 of the C-10 hydroxyl group and the C-11 carbamate (68). The KR domain that is present, predicting
269 the C-11 alcohol function, serves as the substrate for the carbamoyl transferase (*palQ*) in a post-
270 translational modification (69–71). Intriguingly, a domain annotated as a luciferase-like
271 monooxygenase (LLM) initially seems out of place; however, palmerolide A has a hydroxy group
272 at C-10, which is the α-position of the acetate subunit inserted in module 9. LLMs associated with
273 BGCs may not serve as true luciferases, but instead demonstrate oxidizing effects on polyketides
274 and peptides without evidence of corresponding bioluminescence (72, 73). For example, there is
275 an overrepresentation of LLMs in *Candidatus* Entotheonella BGCs without known bioluminescence
276 (74). As demonstrated through individual inactivation of the LLM in the BGC of mensacarin, a Type
277 II PKS system, Msn02, Msn04, and Msn08 have key activity as epoxidases and hydroxlases (73).
278 There are several examples of LLMs in modular Type I PKS systems. OnnC from onnamide and
279 NazB from nazumamide are two LLMs in *Candidatus* Endotheonella that are proposed to serve
280 biosynthetically as hydroxylases (74). In calyculin and mandelalide, the CalD and MndB LLMs
281 catalyze chain shortening reactions through α-hydroxylation and Baeyer-Villiger-type oxidation
282 reactions (61, 63). Phormidolide has a LMM that adds a hydroxy group which is then hypothesized
283 to attack an olefin through a Michael-type addition for cyclization with enzymatic assistance from a
284 pyran synthase (48). The hydroxylation that is key in cyclization of oocydin A is likely installed by
285 OocK or OocM, flavin-dependent monooxygenases that are contiguous to the PKS genes and are
286 thought to act while the substrate is bound to a portion of the PKS megaenzyme (45). It is this
287 hydroxylase activity that we propose for the LLM in module 9. Since the producing bacteria is yet
288 to be cultured, it is not established whether this LLM may also serve a role in bioluminescence
289 and/or quorum sensing. Further evidence for the role of the LLM is provided through alignment
290 against other LLM's. The *pal* BGC contains both LuxR family transcriptional regulator as well as
291 the DNA-binding response regulator. In addition to the annotation within Pfam00296, which
292 includes the bacterial LLMs, the sequence is homologous with the multiple sequence alignments
293 and the hidden Markov models of the TIGR subfamily 04020 of the conserved protein domain family
294 cl19096, which is noted for natural product biosynthesis LLMs (74). The subfamily occurs in both
295 NRPS and PKS systems as well as small proteins with binding of either flavin mononucleotide or
296 coenzyme F420. Alignment of the LLMs from multiple PKS systems, including palmerolide A,
297 shows homology with model sequences from the TIGR subfamily 04020 (**Fig. 4**).
298 The addition of C-5 and C-6 and the reduction of the β-carbonyl to form the C-7 hydroxy group
299 of palmerolide A, is due to module 11, which possesses a KR domain in addition to elongating KS
300 (**Fig. 2**). In the structure of palmerolide A this is followed by the fully reduced subunit from module
301 12 as discussed above. The final elongation results from module 13, which includes DH and KR

domains that contribute to the conjugated ester found as palmerolide A's C-1 through C-3, completing the palmerolide A C_{24} carbon skeleton.

Noncanonical termination condensation domain in *palD* for product cyclization and release. Typically, PKS systems terminate with a thioesterase (TE) domain, leading to release of the polyketide from the megaenzyme (63, 75–77). This canonical domain is not present in the *pal* cluster. Instead, the final module in the *cis*-acting biosynthetic gene cluster includes a truncated condensation domain comprised of 133 amino acid residues; compared to the approximately 450 residues that comprise a standard condensation domain (78) (**Fig. 2**). Condensation domains catalyze cyclization through ester formation in free-standing condensation domains that act in *trans* as well as in NRPS systems (79, 80). In addition, this non-canonical termination domain is not without precedent in hybrid PKS-NRPS and in PKS systems, as basiliskamide and phormidolide both include condensation domains for product release (48, 81). Though the terminal condensation domain in the *pal* BGC is truncated, it maintains much of the HHXXDDG motif (**Fig. 4**), most notably the second histidine, which serves as the catalytic histidine in the condensation reaction (78).

Stereochemical and structural confirmation based on sequence information. KR domains are NADPH-dependent enzymes that belong to the short-chain dehydrogenase superfamily, with Rossman-like folds for co-factor binding (82, 83). Enzymatically, the two KR subtypes, A-Type and B-type, are responsible for stereoselective reduction of β-keto groups and can also determine the stereochemistry of α-substituents. C-type KRs, however, lack reductase activity and often serve as epimerases. A-Type KRs have a key tryptophan residue in the active site, do not possess the LDD amino acid motif, and result in the reduction of β-carbonyls to L-configured hydroxy groups (83). B-Type, which are identified by the presence of an LDD amino acid motif, result in formation of D-configured hydroxy groups (83). The stereochemistry observed in palmerolide A is reflected in the active site sequence information for the D-configured hydroxyl groups from module 5 and module 11 (**Fig. 1** and **Fig. 2**). When an enzymatically active DH domain is within the same module, the stereochemistry of the *cis*- versus *trans*-olefin can be predicted, as the combination of an A-Type KR with a DH results in a *cis*-olefin formation and the combination of a B-Type KR with a DH results in *trans*-olefin formation. The *trans*-α,β-olefins arising from module 7 (C14-C15), module 10 (C8-C9), and module 13 (C2-C3) are each consistent with the presence of a B-Type KR and an active DH. The other three olefins present in the structure of palmerolide A, as noted above, likely have positional and stereochemical influence from E domains (module 4 and module 5) or an ECH domain (module 6).

Additional structural confirmation was obtained through defining the specificity of KS domains using phylogenetic analysis and the *Trans*-AT PKS Polyketide Predictor (TransATor) bioinformatic tool (84). KS domains catalyze the sequential two-carbon elongation steps through a Claisen-like condensation with a resulting β-keto feature (85). Additional domains within a given module can modify the β-carbonyl or add functionality to the adjacent α- or γ-positions (83). Specificity of KSs, based the types of modification located on the upstream acetate subunit were determined and were found to be mostly consistent with the retrobiosynthetic predictions (***SI Appendix*, Fig. S1, *SI Appendix*, Table S1**). For example, the first KS, KS1 (module 1), is predicted to receive a subunit containing a β-branch. KS3 (module 4) and KS4 (module 5) are predicted to receive an upstream monomeric unit with α-methylation and an olefinic shift, consistent with the structure of palmerolide A and with the enzymatic transformations resulting from module 3 and module 4, respectively. Interestingly, the KS associated with the HCS cassette is in a clade of its own in the phylogenetic tree. TransATor also aided in confirming the stereochemical outcomes of the hydroxy groups and olefins, which occur through reduction of the β-carbonyls. The predictions for the D-configured hydroxy groups were consistent with not only the the presence of the LDD motif, indicative of type B-type KR as outlined above, but also with stereochemical determination based on the clades of the KS domains of the receiving modules, KS5 (module 6) and KS11 (module 12). They are also consistent with the structure of palmerolide A. The KS predictions, however, did not aid in confirming reduction of the upstream olefins for KS8 (module 9) and KS12 (module 13).

Additional domains in the intergenic region of the pal BGC. A glycosyl transferase (*palP*) and lactone oxidase (*palH*) that are associated with glycosylation of polyketides are encoded in the palmerolide A BGC (**Fig. 2**), though glycosylated palmerolides have not been isolated. Glycosylation as a means of self-resistance has been described for the macrolide oleandomycin BGC, which encodes not only two glycosyl transferases (one with substrate specificity toward oleandomycin and the other with relaxed specificity), but also a glycosyl hydrolase (86, 87). Glycosylation of the macrolide confers internal resistance to ribosomal inhibition, making the glycosylated macrolide an inactivated pro-drug. Activation occurs by the glycosyl hydrolase which is excreted extracellularly (86, 87). Glycosyl transferases may be a common mechanism to inactivate hydroxylated polyketides in both macrolide producing organisms as well as those non-producers that are seeking resistance (88, 89). The D-arabinono-1,4-lactone oxidase (*palH*) is a FAD-dependent oxidoreductase that may work in concert with the glycosyltransferase. An ATP-binding cassette (ABC) transporter is encoded between the core biosynthetic genes and the genes for the *trans*-acting enzymes. This transporter, which has homology to SryD and contains the key nucleotide-binding domain GGNGSGKS, is one candidate for translocation of the macrolide out of the cell. The other may be a non-canonical macB transport system (Murray et al.) that is not co-linear with the palmerolide BGC. Since it is housed within the BGC, it is likely under the same regulatory control. Additionally, a Band7 protein is encoded. Band7 proteins belong to a ubiquitous family of slipin or stomatin-like integral membrane proteins that are found in all domains of life, yet lack assignment of function (90). Together, these *trans*-acting and intergenic domains decorate the carbon backbone with features that give palmerolide A its unique chemical structure and contribute to its bioactivity. Further work will be required to unravel the details of translocation once either a culture or a genetic system are established.

C. Multiple copies of the *pal* Biosynthetic Gene Cluster Explain Structural Variants in the Palmerolide Family

Careful assembly of the metagenome assembled genome (MAG) of Ca. *S. palmerolidicus* revealed the BGC was present in multiple copies (**Fig. 1** and ***SI Appendix*, Fig. S2-S4**) (40). The structural complexity of the multicopy BGCs represents a biosynthetic system that is similar to that which is found in *Ca*. Didemnitutus mandela, another ascidian-associated verrucomicrobium in the family *Opitutaceae* (63). A total of 5 distinct BGCs consistent with Type-I PKS systems that have much overlap with one another and are likely responsible for much of the structural diversity in the family of palmerolides (17, 18) (**Fig. 1**). Palmerolide A is hypothesized to arise from the BGC designated as *pal* BGC 4 with additional compounds also from this same cluster. The clusters designated as *pal* BGC 1, *pal* BGC 2, *pal* BGC 3, and *pal* BGC 5 have been identified in *Ca.* S. palmerolidicus and the potential products of each are described below. It is hypothesized that there are three levels of diversity introduced to create the family of palmerolides: (1) differences in the site of action for the *trans*-acting domains (with additional *trans*-acting domains at play as well), (2) promiscuity of the initial selection of the starter subunit, and (3) differences in the core biosynthetic genes with additional PKS domains or stereochemical propensities within a module.

There are several palmerolides that likely arise from the same BGC encoding the megaenzyme responsible for palmerolide A (*pal* BGC 4). We hypothesize that the *trans*-acting domains have different sites of action than what is seen in palmerolide A biosynthesis. For example, the chemical scaffold of palmerolide B (**Fig. 1**) is similar to palmerolide A, though the carbamate transfer occurs on the C-7 hydroxy group. Palmerolide B instead bears a sulfate group on the C-11 hydroxy group; proteins with homology to multiple types of sulfatases from the UniProtKB database, (P51691, P15289, O69787, Q8ZQJ2) found in the genome of *Ca.* S. palmerolidicus (40), but are not encoded within the BGCs. One of these *trans*-acting sulfatases likely modifies the molecule post-translationally. Other structural differences including the hydroxylation on C-8 instead of C-10 (as observed in palmerolide A) and the Δ^9 olefin that differs from palmerolide A's

410 Δ^8 olefin, are either due to a difference of the site of action of the LLM (module 9) or a *trans*-acting
411 hydroxylase. Another member of the compound family, palmerolide C, has structural differences
412 attributable to *trans*-acting enzymes as well. Again, a *trans*-acting hydroxylase or the LLM is
413 proposed to be responsible for hydroxylation on C-8. A hydroxy group on C-9 occurs through
414 reduction of the carbonyl. The carbamate installation occurs on C-11 after *trans*-acting
415 hydroxylation or LLM hydroxylation. In addition, the Δ^8 olefin in palmerolide A is not observed, but
416 rather a Δ^6 olefin.
417 Additional levels of structural variation are seen at the site of the starter unit, likely due to
418 a level of enzymatic promiscuity of the second AT (PalF). This, combined with differences in the
419 sites of action for the trans-acting domains, is likely responsible for the structural differences
420 observed in palmerolide F (**Fig. 1**). The terminal olefin on the tail of the macrolide which perhaps
421 is a product of promiscuity of the selection of the starter unit, the isomeric 3-methyl-3-butenoic acid,
422 is consistent with the aforementioned lack of consensus for malonate selection by the AT. In
423 addition, the KS that receives the starter unit is phylogenetically distinct from the other KS in the
424 *pal* clusters (***SI Appendix*, Fig. S1**).
425 The retrobiosynthetic hypothesis for Palmerolide G (**Fig. 1**) has much similarity to what is
426 present in *pal* BGC 4; however, the presence of a *cis*-olefin rather than a *trans*-olefin could arise
427 from a difference in the enzymatic activity of module 4. This olefin is the result of a shift from an
428 epimerase and, therefore, the stereochemistry is not solely reliant upon the action of the associated
429 KR. Although this difference has not been identified in the BGCs in the samples sequenced, this
430 could be present in other environmental samples that have been batched for processing and
431 compound isolation. Currently, the biosynthetic mechanism is unknown.
432 The core biosynthetic genes of two identical palmerolide BGCs (*pal* BGC 1 and *pal* BGC
433 3) (***SI Appendix*, Fig. S2**) possess an additional elongation module when compared to *pal* BGC 4.
434 Palmerolide D (**Fig. 1**) is structurally very similar to palmerolide A with the exception of elongation
435 in the carboxylate tail of the macrolide by an isopropyl group. This could arise from one additional
436 round of starter unit elongation via a KS and methylation. These two identical BGCs are consistent
437 with this structural difference. The overall architecture and stereochemistry are otherwise
438 maintained. Palmerolide H (**Fig. 1**) includes the structural differences of both palmerolide B and
439 palmerolide D. It contains the extended carboxylate tail with a terminal olefin and incorporates
440 hydroxylation on C-8 rather than C-10. Again, there is no genomic evidence that this hydroxylation
441 in the α-position of the acetate subunit is due to incorporation of hydroxy-malonate to explain this
442 but is instead likely due to a *trans*-hydroxylase. The carbamate installation occurs on C-7, while
443 sulfonation occurs on C-11 and α-hydroxy placement is on C-8.
444 The final two palmerolide BGCs both have truncation of the core biosynthetic genes. The
445 gene structure of *pal* BGC 5 (***SI Appendix*, Fig. S3**) shows preservation of genes upstream to the
446 core biosynthetic genes; however, there are no pre-NRP PKS modules noted in the BGC. The HCS
447 cassette, glycosyl transferase, and CT are all present downstream. The predicted product of this
448 cluster does not align with a known palmerolide, though posttranslational hydrolysis of the C-24
449 amide may result in a structure similar to palmerolide E (**Fig. 1**), which maintains much of the
450 structure of palmerolide A; however, is missing the initial polyketide starter unit and the glycine
451 subunit. The final *pal* BGC in *Ca.* S. palmerolidicus, *pal* BGC 2 (***SI Appendix*, Fig. S4**), includes
452 only 5 elongating modules, which would result in a 10-carbon structure that has not been observed.
453 Interestingly, despite the shortened BGC, the HCS cassette, glycosyl transferase, and CT are all
454 present downstream. There would only be a single hydroxy group serving as a substrate for the
455 CT, glycosyl transferase, and sulfatase to act. The 2-carbon site of action for the β-branch
456 introduced in the palmerolide A structures would not be present. The structure-based
457 retrobiosynthesis the eight known palmerolides (A-F) can be hypothesized to arise from differences
458 in the core biosynthetic genes of these non-identical copies of the *pal* BGC, starter unit promiscuity,
459 and differing sites of action in the *trans*-acting enzymes.

460
461
462 **Conclusion**
463
464 The putative *pal* BGC has been identified and represents the first BGC elucidated from an
465 Antarctic organism. As outlined in a retrobiosynthetic strategy, the *pal* BGC represents a *trans*-AT
466 Type I PKS/NRPS hybrid system with compelling alignment to the predicted biosynthetic steps for
467 palmerolide A. The *pal* BGC is proposed to begin with PKS modules resulting in the incorporation
468 of an isovaleric acid derivative, 3-methylcrotonic acid, as a starter unit, followed by incorporation of
469 a glycine residue with NRPS-type modules. Then, eleven rounds of progressive polyketide
470 elongation likely occur leading to varying degrees of oxidation introduced with each module. There
471 are several interesting non-canonical domains encoded within the BGC, such as an HCS, CT, LLM,
472 and a truncated condensation termination domain. Additionally, a glycosylation domain may be
473 responsible for reversible, pro-drug formation to produce self-resistance to the V-ATPase activity
474 of palmerolide A. There are additional domains, the function of which have yet to be determined.
475 A combination of modular alterations, starter unit differences, and activity of *trans*-acting
476 enzymes contributes to Nature's production of a suite of palmerolide analogues. There are a total
477 of five distinct *pal* BGCs in the MAG of *Ca.* S. palmerolidicus, yielding the known eight palmerolides,
478 with genetic differences that explain some of the structural variety seen within this family of
479 compounds. These include differences in modules that comprise the core biosynthetic genes.
480 Additionally, it is proposed that some of the architectural diversity of palmerolides arises from
481 different sites of action of the *trans*-acting, or non-colinear, modules. Starter unit promiscuity is
482 another potential source of the structural differences observed in the compounds. Analysis of the
483 *pal* BGC not only provides insight into the architecture of this Type I PKS/NRPS hybrid BGC with
484 unique features, but also lays the foundational groundwork for drug development studies of
485 palmerolide A via heterologous expression.
486
487
488
489 **Materials and Methods**
490
491 Specimens of *Synoicum adareanum* were collected by SCUBA from the Antarctic Peninsula
492 in the Anvers Island Archipelago and flash frozen. The methods for processing as well as
493 metagenome sequencing and assembly are outlined in detail in two related publications (31, 40).
494 In brief, the first step of data analysis was the result of metagenome sequencing of two Antarctic
495 *S. adareanum* samples (Sample IDs: Nor-2c-2007 and Nor-2a-2007). Ion Proton and 454, two
496 different high throughput DNA sequencing technologies, were utilized. The sequence data was
497 independently assembled using both Newbler (v. 2.9)(91) and SPAdes (v. 3.5) (92). The results
498 were merged via MeGAmerge (v. 1.2) (93). Assembly 1 resulted in 86,387 contigs with a maximum
499 contig size 153,680 and total contig size 144,953,904 bases. The second step in the metagenome
500 sequencing effort was based on two additional *S. adareanum* samples (Bon-1c-2011 and Del-2b-
501 2011) that were also collected by SCUBA and then frozen. These samples were rich in domains
502 for the *pal* BGCs. Six cells of data were obtained from PacBio sequencing, two cells from Bon1C
503 and four cells from Del2B. We generated Circular Consensus Sequences (CCSs) using SMRT
504 Link. The CCS reads were assembled on EDGE Bioinformatics (94) using wtdbg2 (95) to create
505 Assembly 2. The assembled contigs were binned using MaxBin2 (v. 2.2.6) and the binned contigs
506 from different assemblies with 58% GC content were used to re-assemble them via phrap (96), an
507 overlap based assembler. Manual editing of the assembled genome was performed, and CCS
508 reads were used to close gaps and verify repeat regions. We obtained a 4.3 MB genome of
509 palmerolide A producing bacteria in 5 contigs with unique sequences and 5 varying copies of the
510 BGC repeats (referred to as *pal* BGC 1 through *pal* BGC 5).
511 The assembled metagenome was annotated using antiSMASH (v. 5.0). These annotations
512 were integrated with information from protein family homology analysis, conserved domain
513 searches, active site and motif identification to delineate the biosynthetic steps. Manual

interrogations of sequence information of the *pal* BGCs was used to perform BLASTP searches, confirm enzymatic identities, and to identify active site residues key for stereochemical outcomes, substrate affinities, and other granular details. BGCs were obtained from public NCBI databases for basiliskamide, bryostatin 1, calyculin, corallopyronin, mandelalide, onnamide, oxazolamycin, pederin, phormidolide, psymberin, sorangicin, and myxoviricin. ClustalO alignment tool in the CLC Genome Workbench (QIAGEN aarhau A/S v. 20.0.3) was used for multiple sequence alignments of enzymatic domains with the HMM Pfam Seed obtained from EMBL-EBI and the amino acid sequences from the other PKS BGCs. MIBiG (97) was used to acquire the KS amino acid sequence from the type III PKS BGC responsible for 3-(2'-hydroxy-3'-oxo-4'-methylpentyl)-indole biosynthesis from *Xenorhabdus bovienii SS-2004* (GenBank Accession: FN667741.1), which was used for an outgroup. The BGC KSs were numbered according to their position in their proposed biosynthesis in the literature. Prior to the construction of the phylogenetic tree for the KS domains (***SI Appendix*, Fig. S1**), the sequences in the alignment were manually inspected and trimmed. Phylogenetic trees were created in the CLC Genome Workbench (QUIAGEN aarhau A/S v. 20.0.3) with Neighbour Joining (NJ) as a distance method and Bayesian estimation. Jukes-Cantor was selected for the genetic distance model and bootstrapping with performed with 100 replicates. Additionally, the sequence of each KS in the *pal* BGCs was queried using the *Trans*-AT PKS Polyketide Predictor (TransATor) to help define the specificity of KS domains. The software is based on phylogenetic analyses of fifty-four *trans*-AT type I PKS systems with 655 KS sequences and the resulting clades are referenced to help predict the KS specificity for the upstream unit (84).

Data Desposition

Data from this project has been deposited to the NCBI under BioProject Accession Number PRJNA662631. The GenBank Accession Number for the MAG of *Ca.* Synoicohabitans palmerolidicus is JAGGDC000000000. The BGCs can be accessed in the MiBIG database with the associated Accession Numbers: BGC0002118 (for *pal* BGC_4) and BGC0002119 (for *pal* BGC_3).

Acknowledgments

Support for this research was provided in part by the National Institutes of Health award (CA205932) to A.E.M., B.J.B., and P.S.G.C., with additional support from National Science Foundation awards (ANT-0838776, and PLR-1341339 to B.J.B., ANT-0632389 to A.E.M.) The authors acknowledge the assistance of field team members, including William Dent, Charles D. Amsler, James B. McClintock, Margaret O. Amsler, and Katherine Schoenrock. This work would not have been possible without the outstanding logistical support of the United States Antarctic Program. Lucas Bishop, Robert Read, and Mary L. Higham are also recognized for their contributions.

References

1. N. G. Wilson, J. A. Maschek, B. J. Baker, A species flock driven by predation? Secondary metabolites support diversification of slugs in Antarctica. *PLoS One* **8**, 1–7 (2013).
2. C. A. Dejong, *et al.*, Polyketide and nonribosomal peptide retro-biosynthesis and global gene cluster matching. *Nat. Chem. Biol.* **12**, 1007–1014 (2016).
3. J. G. Lawrence, H. Ochman, M. A. Ragan, Reconciling the many faces of lateral gene transfer. *Trends Microbiol.* **10**, 1–4 (2002).
4. M. Ravenhall, N. Škunca, F. Lassalle, C. Dessimoz, Inferring Horizontal Gene Transfer. *PLoS Comput. Biol.* **11**, 1–16 (2015).
5. I. Schmitt, H. T. Lumbsch, Ancient horizontal gene transfer from bacteria enhances

568 biosynthetic capabilities of fungi. *PLoS One* **4**, 1–8 (2009).
569 6. E. W. Schmidt, Trading molecules and tracking targets in symbiotic interactions. *Nat.*
570 *Chem. Biol.* **4**, 466–473 (2008).
571 7. E. W. Schmidt, The secret to a successful relationship: Lasting chemistry between
572 ascidians and their symbiotic bacteria. *Invertebr. Biol.* **134**, 88–102 (2015).
573 8. E. J. N. Helfrich, J. Piel, Biosynthesis of polyketides by *trans*-AT polyketide synthases.
574 *Nat. Prod. Rep.* **33**, 231–316 (2016).
575 9. A. M. P. Koskinen, K. Karisalmi, Polyketide stereotetrads in natural products. *Chem. Soc.*
576 *Rev.* **34**, 677–690 (2005).
577 10. J. Napolitano, A. Daranas, M. Norte, J. Fernandez, Marine Macrolides, a Promising
578 Source of Antitumor Compounds. *Anticancer. Agents Med. Chem.* **9**, 122–137 (2012).
579 11. R. Ueoka, *et al.*, Metabolic and evolutionary origin of actin-binding polyketides from
580 diverse organisms. *Nat. Chem. Biol.* **11**, 705–712 (2015).
581 12. M. E. Bordeleau, *et al.*, Stimulation of mammalian translation initiation factor eIF4A activity
582 by a small molecule inhibitor of eukaryotic translation. *Proc. Natl. Acad. Sci. U. S. A.* **102**,
583 10460–10465 (2005).
584 13. S. Nishimura, *et al.*, Structure-activity relationship study on 13-deoxytedanolide, a highly
585 antitumor macrolide from the marine sponge *Mycale adhaerens*. *Bioorganic Med. Chem.*
586 **13**, 455–462 (2005).
587 14. R. Shen, C. T. Lin, E. J. Bowman, B. J. Bowman, J. A. Porco, Lobatamide C: Total
588 synthesis, stereochemical assignment, preparation of simplified analogues, and V-ATPase
589 inhibition studies. *J. Am. Chem. Soc.* **125**, 7889–7901 (2003).
590 15. E. J. Bowman, K. R. Gustafson, B. J. Bowman, M. R. Boyd, Identification of a New
591 Chondropsin Class of Antitumor Compound that Selectively Inhibits V-ATPases. *J. Biol.*
592 *Chem.* **278**, 44147–44152 (2003).
593 16. S. Kazami, *et al.*, Iejimalides show anti-osteoclast activity via V-ATPase inhibition. *Biosci.*
594 *Biotechnol. Biochem.* **70**, 1364–1370 (2006).
595 17. T. Diyabalanage, C. D. Amsler, J. B. McClintock, B. J. Baker, Palmerolide A, a cytotoxic
596 macrolide from the antarctic tunicate *Synoicum adareanum*. *J. Am. Chem. Soc.* **128**,
597 5630–5631 (2006).
598 18. J. H. Noguez, *et al.*, Palmerolide macrolides from the Antarctic tunicate *Synoicum*
599 *adareanum*. *Bioorganic Med. Chem.* **19**, 6608–6614 (2011).
600 19. B. Shen, *et al.*, Polyketide biosynthesis beyond the type I, II, and III polyketide synthase
601 paradigms: A progress report. *ACS Symp. Ser.* **955**, 154–166 (2007).
602 20. T. A. Nguyen, *et al.*, Exploiting the mosaic structure of *trans*-acyltransferase polyketide
603 synthases for natural product discovery and pathway dissection. *Nat. Biotechnol.* **26**, 225–
604 233 (2008).
605 21. A. Nivina, K. P. Yuet, J. Hsu, C. Khosla, Evolution and Diversity of Assembly-Line
606 Polyketide Synthases. *Chem. Rev.* **119**, 12524–12547 (2019).
607 22. P. Videau, K. N. Wells, A. J. Singh, W. H. Gerwick, B. Philmus, Assessment of *Anabaena*
608 sp. Strain PCC 7120 as a Heterologous Expression Host for Cyanobacterial Natural
609 Products: Production of Lyngbyatoxin A. *ACS Synth. Biol.* **5**, 978–988 (2016).
610 23. E. J. Kim, *et al.*, Heterologous Production of 4-O-Demethylbarbamide, a Marine
611 Cyanobacterial Natural Product. *Org. Lett.* **14**, 5824–5827 (2012).
612 24. C. Greunke, *et al.*, Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic
613 potential. *Metab. Eng.* **47**, 334–345 (2018).
614 25. J. B. McClintock, C. D. Amsler, B. J. Baker, R. W. M. Van Soest, Ecology of antarctic
615 marine sponges: An overview. *Integr. Comp. Biol.* **45**, 359–368 (2005).
616 26. A. Clarke, J. A. Crame, Evolutionary dynamics at high latitudes: Speciation and extinction
617 in polar marine faunas. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 3655–3666 (2010).
618 27. R. M. Young, *et al.*, Site-specific variability in the chemical diversity of the Antarctic red
619 alga Plocamium cartilagineum. *Mar. Drugs* **11**, 2126–2139 (2013).
620 28. K. D. Paull, E. Hamel, L. Malspeis, Prediction of Biochemical Mechanism of Action from
621 the In Vitro Antitumor Screen of the National Cancer Institute. *Cancer Chemother. Agents*,

9–45 (1995).
29. K. Von Schwarzenberg, *et al.*, Mode of cell death induction by pharmacological vacuolar H +-ATPase (V-ATPase) inhibition. *J. Biol. Chem.* **288**, 1385–1396 (2013).
30. T. Diyabalanage, Chemical Investigation of two Antarctic Invertebrates, Synoicum adareanum (Chordata: Ascidiaceae; Enterogona; Polyclinidae) and *Austrodoris kergulenensis* (Molusca; Gastropoda; Nudibranchia; Dorididae). 211 (2006).
31. A. E. Murray, *et al.*, Uncovering the Core Microbiome and Distribution of Palmerolide in *Synoicum adareanum* Across the Anvers Island Archipelago, Antarctica. *Mar. Drugs* **18**, 298 (2020).
32. K. P. Kaliappan, P. Gowrisankar, Synthetic studies on a marine natural product, palmerolide A: Synthesis of C1-C9 and C15-C21 fragments. *Synlett*, 1537–1540 (2007).
33. M. P. Lisboa, D. M. Jones, G. B. Dudley, Formal synthesis of palmerolide A, featuring alkynogenic fragmentation and syn-selective vinylogous aldol chemistry. *Org. Lett.* **15**, 886–889 (2013).
34. M. D. Lebar, B. J. Baker, Synthesis of the C3-14 fragment of palmerolide A using a chiral pool based strategy. *Tetrahedron* **66**, 1557–1562 (2010).
35. Z. K. Wen, Y. H. Xu, T. P. Loh, Palladium-catalyzed cross-coupling of unactivated alkenes with acrylates: Application to the synthesis of the C13-C21 fragment of palmerolide A. *Chem. - A Eur. J.* **18**, 13284–13287 (2012).
36. S. A. Pujari, P. Gowrisankar, K. P. Kaliappan, A shimizu non-aldol approach to the formal total synthesis of palmerolide A. *Chem. - An Asian J.* **6**, 3137–3151 (2011).
37. K. R. Prasad, A. B. Pawar, Stereoselective synthesis of C1-C18 region of palmerolide A from tartaric acid. *Synlett* **4**, 1093–1095 (2010).
38. A. B. Pawar, K. R. Prasad, Formal total synthesis of palmerolide A. *Chem. - A Eur. J.* **18**, 15202–15206 (2012).
39. C. S. Riesenfeld, A. E. Murray, B. J. Baker, Characterization of the microbial community and polyketide biosynthetic potential in the palmerolide-producing tunicate *Synoicum adareanum*. *J. Nat. Prod.* **71**, 1812–1818 (2008).
40. A. E. Murray, *et al.*, Discovery of an Antarctic ascidian-associated uncultivated verrucomicrobium that encodes the biosynthetic capability for antimelanoma palmerolides. **(In prep.)** (2021).
41. W. Mattheus, *et al.*, Isolation and Purification of a New Kalimantacin/Batumin-Related Polyketide Antibiotic and Elucidation of Its Biosynthesis Gene Cluster. *Chem. Biol.* **17**, 149–159 (2010).
42. S. Sudek, *et al.*, Identification of the putative bryostatin polyketide synthase gene cluster from “Candidatus Endobugula sertula”, the uncultivated microbial symbiont of the marine bryozoan *Bugula neritina*. *J. Nat. Prod.* **70**, 67–74 (2007).
43. D. J. Edwards, *et al.*, Structure and Biosynthesis of the Jamaicamides, New Mixed Polyketide-Peptide Neurotoxins from the Marine Cyanobacterium *Lyngbya majuscula*. *Chem. Biol.* **11**, 817–833 (2004).
44. Z. Chang, *et al.*, Biosynthetic Pathway and Gene Cluster Analysis of Curacin A, an Antitubulin Natural Product from the Tropical Marine Cyanobacterium *Lyngbya m ajuscula* †. *J. Nat. Prod.* **67**, 1356–1367 (2004).
45. M. A. Matilla, H. Stöckmann, F. J. Leeper, G. P. C. Salmond, Bacterial biosynthetic gene clusters encoding the anti-cancer haterumalide class of molecules: Biogenesis of the broad spectrum antifungal and anti-oomycete compound, oocydin A. *J. Biol. Chem.* **287**, 39125–39138 (2012).
46. K. M. Fisch, *et al.*, Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. *Nat. Chem. Biol.* **5**, 494–501 (2009).
47. T. A. M. Gulder, M. F. Freeman, J. Piel, “The Catalytic Diversity of Multimodular Polyketide Synthases: Natural Product Biosynthesis Beyond Textbook Assembly Rules” in (2011) https:/doi.org/10.1007/128_2010_113.
48. M. J. Bertin, *et al.*, The Phormidolide Biosynthetic Gene Cluster: A trans -AT PKS Pathway Encoding a Toxic Macrocyclic Polyketide. *ChemBioChem* **17**, 164–173 (2016).

676 49. C. D. Reeves, *et al.*, Alteration of the substrate specificity of a modular polyketide
677 synthase acyltransferase domain through site-specific mutations. *Biochemistry* **40**, 15464–
678 15470 (2001).
679 50. G. Yadav, R. S. Gokhale, D. Mohanty, Computational approach for prediction of domain
680 organization and substrate specificity of modular polyketide synthases. *J. Mol. Biol.* **328**,
681 335–363 (2003).
682 51. B. J. Dunn, K. R. Watts, T. Robbins, D. E. Cane, C. Khosla, Comparative analysis of the
683 substrate specificity of *trans* - Versus *cis*- acyltransferases of assembly line polyketide
684 synthases. *Biochemistry* **53**, 3796–3806 (2014).
685 52. A. L. Díaz-Pérez, C. Díaz-Pérez, J. Campos-García, Bacterial l-leucine catabolism as a
686 source of secondary metabolites. *Rev. Environ. Sci. Bio/Technology* **15**, 1–29 (2016).
687 53. T. Mahmud, *et al.*, A novel biosynthetic pathway providing precursors for fatty acid
688 biosynthesis and secondary metabolite formation in myxobacteria. *J. Biol. Chem.* **277**,
689 32768–32774 (2002).
690 54. M. Röttig, *et al.*, NRPSpredictor2 - A web server for predicting NRPS adenylation domain
691 specificity. *Nucleic Acids Res.* **39**, 362–367 (2011).
692 55. K. Blin, *et al.*, antiSMASH 5.0: updates to the secondary metabolite genome mining
693 pipeline. *Nucleic Acids Res.* **47**, W81–W87 (2019).
694 56. D. C. Cantu, Y. Chen, P. J. Reilly, Thioesterases: A new perspective based on their
695 primary and tertiary structures. *Protein Sci.* **19**, 1281–1295 (2010).
696 57. J. Moldenhauer, *et al.*, The Final Steps of Bacillaene Biosynthesis in *Bacillus*
697 *amyloliquefacien*s FZB42: Direct Evidence for β,γ Dehydration by a *trans*-Acyltransferase
698 Polyketide Synthase. *Angew. Chemie* **122**, 1507–1509 (2010).
699 58. G. Berkhan, C. Merten, C. Holec, F. Hahn, The Interplay between a Multifunctional
700 Dehydratase Domain and a C-Methyltransferase Effects Olefin Shift in Ambruticin
701 Biosynthesis. *Angew. Chemie - Int. Ed.* **55**, 13589–13592 (2016).
702 59. C. Hertweck, The biosynthetic logic of polyketide diversity. *Angew. Chemie - Int. Ed.* **48**,
703 4688–4716 (2009).
704 60. J. Moldenhauer, X. H. Chen, R. Borriss, J. Piel, Biosynthesis of the antibiotic bacillaene,
705 the product of a giant polyketide synthase complex of the *trans*-AT family. *Angew. Chemie*
706 *- Int. Ed.* **46**, 8195–8197 (2007).
707 61. T. Wakimoto, *et al.*, Calyculin biogenesis from a pyrophosphate protoxin produced by a
708 sponge symbiont. *Nat. Chem. Biol.* **10**, 648–655 (2014).
709 62. E. Esquenazi, *et al.*, Visualizing the spatial distribution of secondary metabolites produced
710 by marine cyanobacteria and sponges via MALDI-TOF imaging. *Mol. Biosyst.* **4**, 562–570
711 (2008).
712 63. J. Lopera, I. J. Miller, K. L. McPhail, J. C. Kwan, Increased Biosynthetic Gene Dosage in a
713 Genome-Reduced Defensive Bacterial Symbiont. *mSystems* **2** (2017).
714 64. T. J. Buchholz, *et al.*, Polyketide β-Branching in Bryostatin Biosynthesis: Identification of
715 Surrogate Acetyl-ACP Donors for BryR, an HMG-ACP Synthase. *Chem. Biol.* **17**, 1092–
716 1100 (2010).
717 65. H. Irschik, *et al.*, Analysis of the sorangicin gene cluster reinforces the utility of a combined
718 phylogenetic/retrobiosynthetic analysis for deciphering natural product assembly by *trans*-
719 AT PKS. *ChemBioChem* **11**, 1840–1849 (2010).
720 66. Ö. Erol, *et al.*, Biosynthesis of the myxobacterial antibiotic corallopyronin A.
721 *ChemBioChem* **11**, 1253–1265 (2010).
722 67. Y. Q. Cheng, G. L. Tang, B. Shen, Type I polyketide synthase requiring a discrete
723 acyltransferase for polyketide biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3149–
724 3154 (2003).
725 68. K. C. Nicolaou, *et al.*, Chemical synthesis and biological evaluation of palmerolide A
726 analogues. *J. Am. Chem. Soc.* **130**, 10019–10023 (2008).
727 69. S. F. Haydock, *et al.*, Organization of the biosynthetic gene cluster for the macrolide
728 concanamycin A in Streptomyces neyagawaensis ATCC 27449. *Microbiology* **151**, 3161–
729 3169 (2005).

730 70. W. Chen, *et al.*, Characterization of the Polyoxin Biosynthetic Gene Cluster from
731 *Streptomyces cacaoi* and Engineered Production of Polyoxin H. *J. Biol. Chem.* **284**,
732 10627–10638 (2009).
733 71. T. K. Mihali, W. W. Carmichael, B. A. Neilan, A putative gene cluster from a *Lyngbya*
734 *wollei* bloom that encodes paralytic shellfish toxin biosynthesis. *PLoS One* **6** (2011).
735 72. A. K. El-Sayed, J. Hothersall, C. M. Thomas, Quorum-sensing-dependent regulation of
736 biosynthesis of the polyketide antiobiotic mupirocin in *Pseudomonas fluorescens* NCIMB
737 10586. *Microbiology* **147**, 2127–2139 (2001).
738 73. S. Maier, *et al.*, Functional characterization of different ORFs including luciferase-like
739 monooxygenase genes from the mensacarcin gene cluster. *ChemBioChem* **16**, 1175–
740 1182 (2015).
741 74. G. Lackner, E. E. Peters, E. J. N. Helfrich, J. Piel, Insights into the lifestyle of uncultured
742 bacterial natural product factories associated with marine sponges. *Proc. Natl. Acad. Sci.*
743 **114**, E347–E356 (2017).
744 75. J. J. Gehret, *et al.*, Terminal alkene formation by the thioesterase of curacin A
745 biosynthesis: Structure of a decarboxylating thioesterase. *J. Biol. Chem.* **286**, 14445–
746 14454 (2011).
747 76. L. Gu, *et al.*, Metamorphic enzyme assembly in polyketide diversification. *Nature* **459**,
748 731–735 (2009).
749 77. J. Piel, A polyketide synthase-peptide synthetase gene cluster from an uncultured
750 bacterial symbiont of Paederus beetles. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 14002–14007
751 (2002).
752 78. T. Stachelhaus, H. D. Mootz, V. Bergendahl, M. A. Marahiel, Peptide Bond Formation in
753 Nonribosomal Peptide Biosynthesis. *J. Biol. Chem.* **273**, 22773–22781 (1998).
754 79. S. Lin, S. G. Van Lanen, B. Shen, A free-standing condensation enzyme catalyzing ester
755 bond formation in C-1027 biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 4183–4188
756 (2009).
757 80. K. Zaleta-Rivera, *et al.*, A bidomain nonribosomal peptide synthetase encoded by FUM14
758 catalyzes the formation of tricarballylic esters in the biosynthesis of fumonisins.
759 *Biochemistry* **45**, 2561–2569 (2006).
760 81. C. M. Theodore, *et al.*, Genomic and metabolomic insights into the natural product
761 biosynthetic diversity of a feral-hog-associated *Brevibacillus laterosporus* strain. *PLoS*
762 *One* **9**, 3–12 (2014).
763 82. A. T. Keatinge-Clay, R. M. Stroud, The Structure of a Ketoreductase Determines the
764 Organization of the β-Carbon Processing Enzymes of Modular Polyketide Synthases.
765 *Structure* **14**, 737–748 (2006).
766 83. A. T. Keatinge-Clay, The structures of type I polyketide synthases. *Nat. Prod. Rep.* **29**,
767 1050 (2012).
768 84. E. J. N. Helfrich, *et al.*, Automated structure prediction of *trans*-acyltransferase polyketide
769 synthase products. *Nat. Chem. Biol.* **15**, 813–821 (2019).
770 85. C. Khosla, Y. Tang, A. Y. Chen, N. A. Schnarr, D. E. Cane, Structure and Mechanism of
771 the 6-Deoxyerythronolide B Synthase. *Annu. Rev. Biochem.* **76**, 195–221 (2007).
772 86. L. M. Quirós, I. Aguirrezabalaga, C. Olano, C. Méndez, J. A. Salas, Two
773 glycosyltransferases and a glycosidase are involved in oleandomycin modification during
774 its biosynthesis by Streptomyces antibioticus. *Mol. Microbiol.* **28**, 1177–1185 (1998).
775 87. T. A. Wencewicz, Crossroads of Antibiotic Resistance and Biosynthesis. *J. Mol. Biol.* **431**,
776 3370–3399 (2019).
777 88. G. Jenkins, E. Cundliffe, Cloning and characterization of two genes from Streptomyces
778 lividans that confer inducible resistance to lincomycin and macrolide antibiotics. *Gene* **108**,
779 55–62 (1991).
780 89. A. Gourmelen, M. H. Blondelet-Rouault, J. L. Pernodet, Characterization of a glycosyl
781 transferase inactivating macrolides, encoded by gimA from *Streptomyces ambofaciens*.
782 *Antimicrob. Agents Chemother.* **42**, 2612–2619 (1998).
783 90. M. Boehm, *et al.*, Structural and mutational analysis of band 7 proteins in the

784 cyanobacterium *Synechocystis* sp. strain PCC 6803. *J. Bacteriol.* **191**, 6425–6435 (2009).
785 91. M. J. Chaisson, P. A. Pevzner, Short read fragment assembly of bacterial genomes.
786 *Genome Res.* **18**, 324–330 (2008).
787 92. S. Nurk, *et al.*, Assembling single-cell genomes and mini-metagenomes from chimeric
788 MDA products. *J. Comput. Biol.* **20**, 714–737 (2013).
789 93. M. Scholz, C. C. Lo, P. S. G. Chain, Improved assemblies using a source-agnostic
790 pipeline for metagenomic assembly by merging (MeGAMerge) of contigs. *Sci. Rep.* **4**, 1–9
791 (2014).
792 94. P. E. Li, *et al.*, Enabling the democratization of the genomics revolution with a fully
793 integrated web-based bioinformatics platform. *Nucleic Acids Res.* **45**, 67–80 (2017).
794 95. J. Ruan, H. Li, Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* **17**, 155–
795 158 (2020).
796 96. B. Ewing, P. Green, Base-calling of automated sequencer traces using phred. II. Error
797 probabilities. *Genome Res.* **8**, 186–194 (1998).
798 97. S. A. Kautsar, *et al.*, MIBiG 2.0: A repository for biosynthetic gene clusters of known
799 function. *Nucleic Acids Res.* **48**, D454–D458 (2020).
800

Figures and Tables

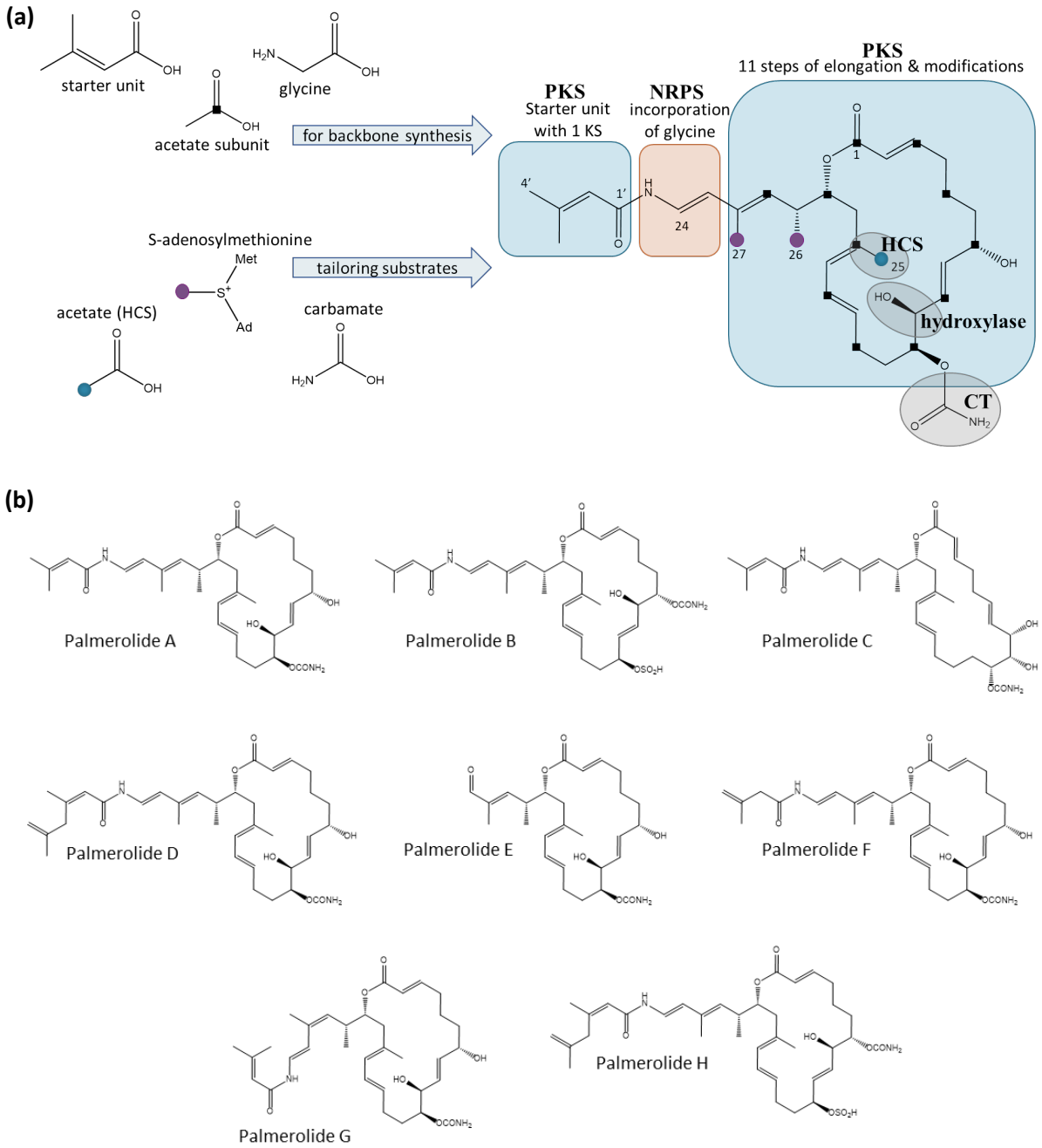

Fig. 1. (a) Structure of palmerolide A with notations for the proposed retrobiosynthesis. Backbone synthesis is a result of incorporation of the starter unit, a glycine residue, and acetate subunits (C1 indicated by black squares). Structural features from *trans*-acting tailoring enzymes (indicated by grey ovals) utilize additional substrates: methyl transfers from SAM (purple dots), installation of C-25 methyl from acetate (blue dot) via HCS cassette, and carbamoyl transfer to the secondary alcohol on C-11. The α-hydroxy group on C-10 from incorporation of hydroxymalonic acid or a *trans*-acting hydroxylase. **(b)** Family of palmerolides. Much of the structural diversity can be

813 explained by differences due to starter unit promiscuity, sites of action for the *trans*-acting
814 tailoring enzymes, and differences in the core modules of the multiple pal BGCs.
815
816

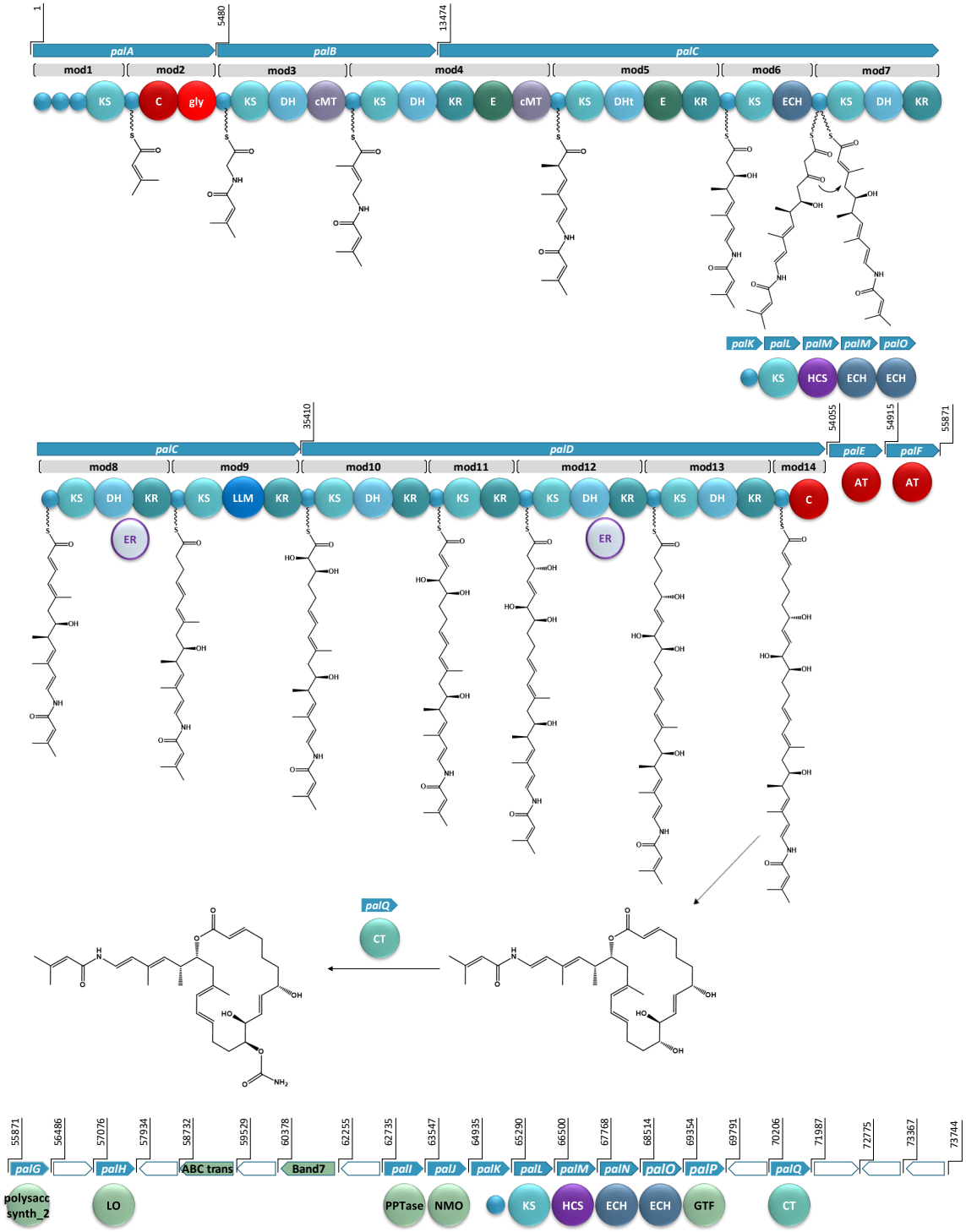

Fig. 2. The proposed BGC for palmerolide A, showing the hybrid PKS-NRPS system. KS: ketosynthase domain, C: condensation domain, gly: adenylation domain for glycine incorporation, DH: dehydratase domain, cMT: carbon methyl transferase domain, KR: ketoreductase domain, E: epimerization domain, DHt: dehydratase variant; ECH: enoyl-CoA hydratase, LLM: luciferase-like monooxygenase, AT: acyl transferase; polysacc synt_2: polysaccharide biosynthesis protein, LO:

823 lactone oxidase, ABC trans: ATP-binding cassette transporter, Band7: stomatin-like integral
824 membrane, PPTase: phosphopantetheinyl transferase, NMO: nitronate monooxygenase, HCS:
825 hydroxymethylglutaryl-CoA synthase, GTF: glycosyl transferase ER: enoyl reductase, CT:
826 carbamoyl transferase, small blue circles represent acyl- or peptidyl-carrier proteins. Blue arrows
827 indicate biosynthetic genes. White arrows reflect hypothetical genes. The BGC is displayed in
828 reverse compliment.
829
830

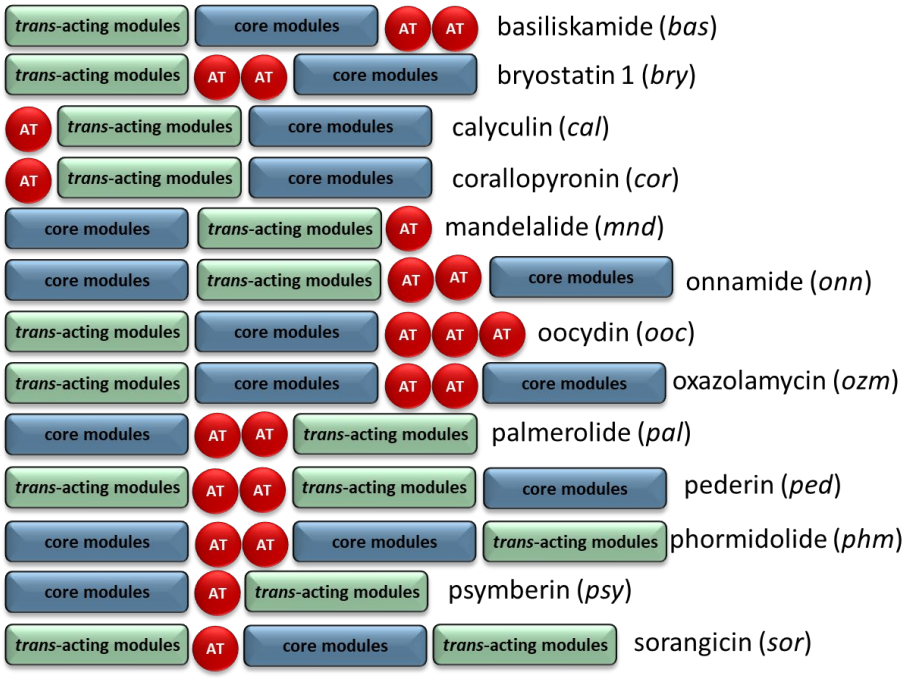

Fig. 3. Comparison of BGC organization of select *trans*-AT systems. There is significant variability in the order of the core modules, AT modules, and modules which contain *trans*-acting tailoring enzymes. There is also variability in the number of encoded AT modules, though the AT modules are typically encoded on separate, but tandem genes if more than one is present.

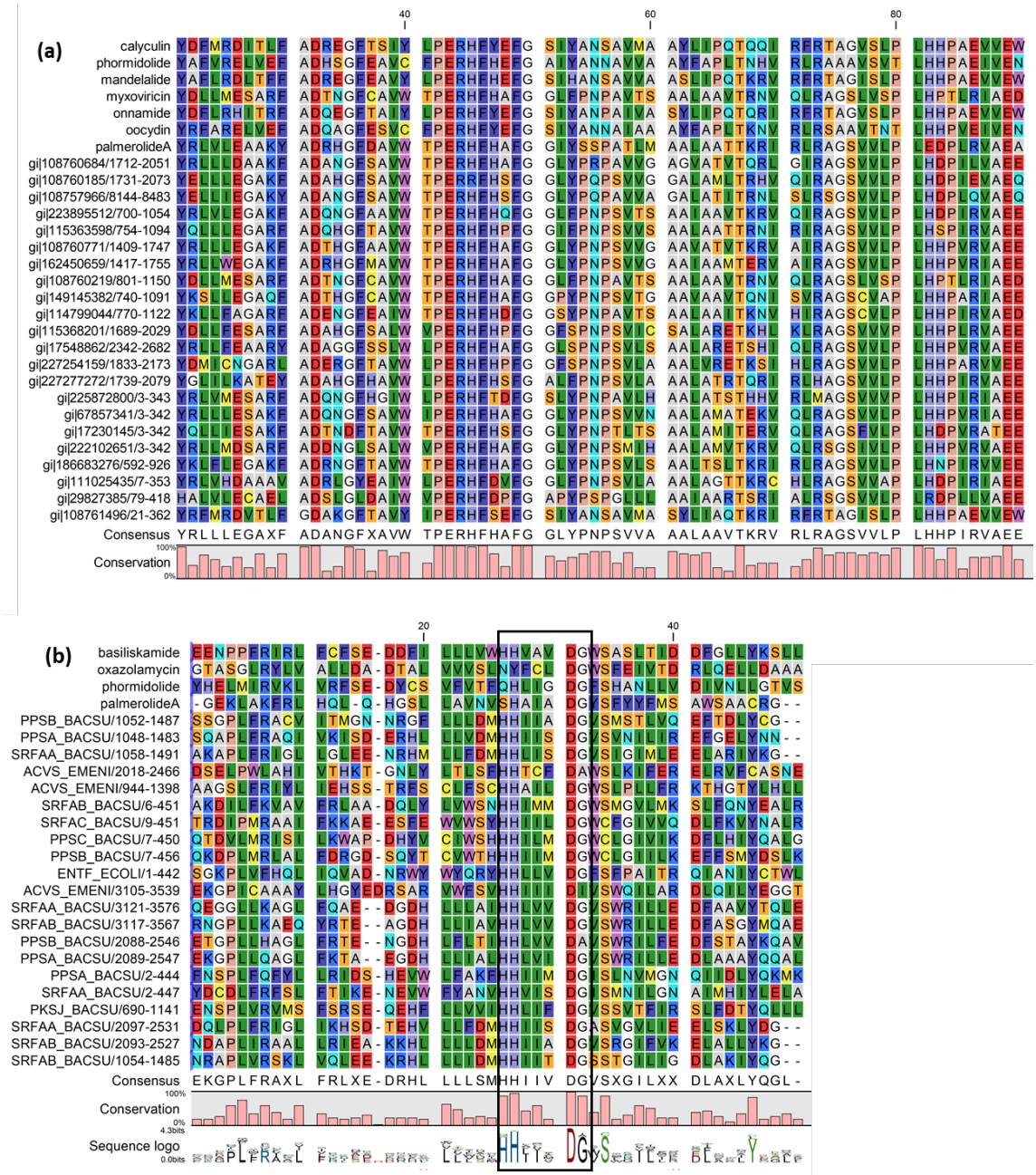

Fig. 4. (a) Alignment of the proposed palmerolide hydroxylase luciferase-like monooxygenase (LLM) with sequences from the TIGR subfamily 04020. **(b)** Alignment of the proposed termination condensation domain of the *pal* BGC_4 with that of basiliskamide and phormidolide as well as the HMM seed sequences for PF00668 (condensation domains); conservation of the HHXXXDG motif can be seen.