Bioinformatic and mechanistic analysis of the palmerolide PKS-NRPS biosynthetic pathway 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, USA

[2]Division of Earth and Ecosystem Sciences, Desert Research Institute, Reno, NV, USA

[3]Los Alamos National Laboratory, Los Alamos, NM, USA

***Corresponding Authors**

Bill J. Baker
bjbaker@usf.edu

Alison E. Murray
alison.murray@dri.edu

Patrick S. G. Chain
pchain@lanl.gov

Keywords

marine natural products, macrolide, biosynthetic gene clusters, Antarctic microbiology, *trans*-AT Type I polyketide synthase, secondary metabolites

Abstract

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 host-associated microbes. These unique metabolites often have bioactivity targets in human disease and can be purposed as pharmaceuticals. Polyketides are a complex family of natural products that often serve as defensive metabolites for competitive or pro-survival purposes for the producing organism, while demonstrating bioactivity in human diseases as cholesterol lowering agents, anti-infectives, and 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 (BGCs) responsible for production of secondary metabolites are encoded in the genomes of the producers as discrete genomic elements. A candidate palmerolide BGC was identified from a *S. adareanum* microbiome-metagenome based on a high degree of congruence with a chemical structure-based retrobiosynthetic prediction. Protein family homology analysis, conserved domain searches, active site and motif identification were used to identify and propose the function of the ~75 kb *trans*-acyltransferase (AT) polyketide synthase-non-ribosomal synthase (PKS-NRPS) domains responsible for the stepwise 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 is found in multiple copies within the metagenome-assembled genome (MAG). 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 microbial species, recently proposed as *Candidatus* Synoicohabitans palmerolidicus. These findings have relevance for fundamental knowledge of PKS combinatorial biosynthesis and could enhance drug development efforts of palmerolide A through heterologous gene expression.

Introduction

Marine invertebrates such as corals, sponges, mollusks, and ascidians, are known to be a rich source of bioactive compounds (Carroll et al., 2019). 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 synthetase (NRPS) products, ribosomally synthesized and post-translationally modified peptides (RiPPs), and polyketides. It is estimated that over 11,000 secondary metabolites from marine and terrestrial environments understood to be products of polyketide synthase (PKS) and NRPS origin have been isolated and described (Dejong et al., 2016). BGCs exist as a series of genomic elements that encode for the biosynthetic machinery 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 (Lawrence et al., 2002; Ravenhall et al., 2015), suggesting a mechanism of horizontal gene transfer from organisms that are distantly related, including across different kingdoms (Schmidt, 2008; Schmitt and Lumbsch, 2009). 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 (Schmidt, 2015).

Polyketides are a complex family of natural products produced by a variety of PKS enzymes that are related to, but evolutionarily divergent from, fatty acid synthases (Helfrich and Piel, 2016). 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 (Koskinen and Karisalmi, 2005). This potential for use as pharmaceuticals is approximately five times greater than for compounds of all other natural product classes (Koskinen and Karisalmi, 2005). Many polyketides are classified as macrolides, which are large-ring lactones that are pharmaceutically relevant due to a number of biological actions, including, targeting the cytoskeleton, ribosomal protein biosynthesis, and vacuolar type V-ATPases (Bordeleau et al., 2005; Nishimura et al., 2005; Napolitano et al., 2012; Ueoka et al., 2015). 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 metastasis (Napolitano et al., 2012). A number of marine macrolides inhibit V-ATPases, including lobatamides, chondropsins, iejimalides, and several of the palmerolides (Bowman et al., 2003; Shen et al., 2003; Diyabalanage et al., 2006; Kazami et al., 2006; Noguez et al., 2011).

There are three types of PKS systems. Type I PKS systems in bacteria are primarily comprised of non-iteratively acting multimodular enzymes that lead to progressive elongation of a polyketide chain, though these megaenzymes can also include “stuttering” modules that may act iteratively (Wilkinson et al., 2000; Shen et al., 2007; Tatsuno et al., 2007). In addition, some bacterial Type I PKS systems are comprised solely of iteratively acting monomodular enzymes that catalyze a series of chain elongation steps for polyketide formation (Wang et al., 2020). Type II PKS systems typically contain separate, iteratively acting enzymes that biosynthesize polycyclic aromatic polyketides, while Type III PKS systems possess iteratively-acting homodimeric enzymes that often result in monocyclic or bicyclic aromatic polyketides (Shen et al., 2007). Type I PKS systems can be subdivided into two groups, depending upon whether the acyl transferase (AT) modules are encoded within each module at the site that is parallel to the functional role of the ATs, referred to as *cis*-AT Type I PKS, or physically distinct from the megaenzyme, which are referred to as *trans*-AT Type I PKS. In both cases, there are often parallel relationships between the genome order, the action of enzymatic modules, and the functional groups present in the growing polyketide chain, though in *trans*-AT systems deviations from these parallel relationships is more likely to be observed (Nguyen et al., 2008). In *trans*-AT systems, AT domains may be incorporated in a mosaic fashion through horizontal gene transfer (Nguyen et al., 2008). 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 (Haydock et al., 1995; Jenke-Kodama et al., 2005). Additionally, recombination, gene duplication, and conversion events can lead to further diversification of the resultant biosynthetic machinery (Nivina et al., 2019). 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 (Kim et al., 2012; Videau et al., 2016; Greunke et al., 2018).

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 (McClintock et al., 2005). 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 (McClintock et al., 2005; Clarke and Crame, 2010; Young et al., 2013).

Palmerolide A (**Figure 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 (Diyabalanage et al., 2006). Palmerolide A is a macrolide polyketide that possesses potent bioactivity against malignant melanoma cell lines, while demonstrating minimal cytotoxicity against other cell lines (Diyabalanage et al., 2006). The National Cancer Institute’s COMPARE algorithm was used to correlate experimental findings with a database for prediction of the biochemical mechanism of action by identifying the mechanism of action of palmerolide A as a V-ATPase inhibitor (Paull et al., 1995). Downstream effects of V-ATPase inhibition include an increase in both hypoxia induction factor-1α and autophagy (Diyabalanage et al., 2006; Von Schwarzenberg et al., 2013). Increased expression of V-ATPase on the surface of metastatic melanoma cells (Von Schwarzenberg et al., 2013) perhaps explain palmerolide A’s selectivity for UACC-62 cell lines over all cell types (Diyabalanage et al., 2006). 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) (Murray et al., 2020), isolation of palmerolide A from its Antarctic source in mass sufficient for drug development it is neither ecologically nor logistically feasible. Although synthetic strategies for palmerolide A have been reported (Jiang et al., 2007; Kaliappan and Gowrisankar, 2007; Nicolaou et al., 2008b; Penner et al., 2009; Lebar and Baker, 2010; Pujari et al., 2011; Pawar and Prasad, 2012; Lisboa et al., 2013), a clear pathway to achieve sufficient quantities needed for drug development has been elusive. Therefore, there is substantial

interest in identifying the BGC responsible for palmerolide A production as this would pave a way for future drug development efforts.

Our approach to identify the palmerolide BGC (*pal* BGC) began with the characterization of the ascidian host-associated microbiome (Riesenfeld et al., 2008). Next, a persistent cohort of bacteria present across many individual ascidians – a core microbiome – for *Synoicum adareanum* was identified through analysis of occurrence of distinct amplicon sequence variants (ASV) from iTag sequencing of the Variable 3-4 regions of the bacterial 16S rRNA (Murray et al., 2020). This work ultimately led to the evaluation of the microbiome metagenome and the subsequent assembly of a metagenome assembled genome (MAG) of *Candidatus* Synoicohabitans palmerolidicus, a verrucomicrobium in the family *Opitutaceae* (Murray et al., 2021). Contained within the genome are five non-identical copies of a candidate *pal* BGC. Here, we report on a detailed bioinformatic analysis of the *pal* BGCs and conclude that at least three of the candidate BGCs likely are responsible for the biosynthesis of palmerolides with structures that have been previously reported from Antarctic *S. adareanum* in this macrolide family (Diyabalanage et al., 2006; Noguez et al., 2011).

Materials and Methods

The details of field collections, sample processing and genomic methods (sequencing, assembly, and analysis) used to identify the putative biosynthetic gene clusters are described in detail in two related publications (Murray et al., 2020, 2021). In brief, *Synoicum adareanum* samples were collected by SCUBA from the Antarctic Peninsula in the Anvers Island Archipelago and flash frozen. Microbial cells were separated from host tissue using a homogenization protocol established for this holobiont followed by differential centrifugation to separate the host cellular debris and microbial cells (Murray et al., 2020). The metagenome assembled genome (MAG) sequence analyzed in this study was generated from two *S. adareanum* samples (Bon-1c-2011 and Del-2b-2011) with high copy numbers of the putative biosynthetic gene cluster sequenced using PacBio technology. The resulting assembly of produced a nearly complete 4.3 MB genome, with five unique contigs and five varying copies of the *pal* BGC (referred to as *pal* BGC 1 through *pal* BGC 5) (Murray et al., 2021). Given the phylogenomic novelty of the MAG, the name *Candidatus* Synoicihabitans palmerolidicus was proposed for this new genus in the *Opitutaceae* family (Verrucomicrobia phylum).

The methods employed in this study used bioinformatic tools to develop predictive models of palmerolide biosynthesis. Enzymatic reactions and organic synthetic interpretations were based on homology analyses. Automated annotation and manual bioinformatic tools were used to discern the details of palmerolide A biosynthesis in addition to generating predictions for the other BGCs. The *Ca*. S. palmerolidicus MAG was annotated using antiSMASH (v. 5.0) (Blin et al., 2019) using the full complement of annotation options available. Then we predicted the gene cluster responsible for palmerolide A biosynthesis using retrobiosynthetic predictions focused on the 5' end of the BGCs (**Figure 1**). Only one of the five BGCs met the criteria for a non-ribosomal peptide glycine starter unit. The annotation predictions were integrated and validated with results of additional protein family homology analysis, conserved domain searches, active site and motif identification to predict the step-wise biosynthesis of palmerolide A. Manual annotation of the *pal* BGC sequences included BLASTP searches to confirm enzymatic identities, then protein family alignments were used to identify active site residues key for stereochemical outcomes, confirm substrate affinities, and other biochemical synthesis details.

Additional manual bioinformatic efforts included obtaining BGCs from public NCBI databases for basiliskamide, bryostatin 1, calyculin, corallopyronin, mandelalide, onnamide, oxazolamycin, pederin, phormidolide, psymberin, sorangicin, and myxoviricin (**Supplemental Table S2**). 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 HMM Pfam Seeds obtained from EMBL-EBI and the amino acid sequences from the other PKS BGCs. MIBiG (Kautsar et al., 2020) 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 *pal* BGC ACPs and PCPs were numbered according to their position in the proposed biosynthesis of palmerolide A (**Supplemental Figures S1 and S2**). 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 (**Supplemental Figure S4**), the sequences in the alignment were manually inspected and trimmed. Phylogenetic trees were created in the CLC Genome Workbench (QIAGEN aarhau A/S v. 20.0.3) with Neighbour Joining (NJ) as a distance method and Bayesian estimation for ACP and PCP comparisons as well as for KS analysis. 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 (*trans*ATor) 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 (Helfrich et al., 2019).

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 (**Figure 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, incorporation of methylmalonyl CoA or enzymatic activity of carbon methyltransferases (cMTs) were predicted to be responsible for 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. For example, as seen in the kalimantacins (Mattheus et al., 2010), the carbamate on C-11 was hypothesized to be installed by a carbamoyl transferase (CT). The C-25 methyl group located on C-17 in the β-position to the carbonyl suggests the origin of this branch is likely from hydroxymethylglutaryl-CoA synthase (HCS) catalysis, rather than SAM-mediated methylation, which occurs at the α-position to the carbonyl. SAM-mediated methylation does however appear to be the origin of the C-26 and C-27 methyl groups. Methylation in this acetate carboxyl position is unusual, but represented in a number of notable BGC's, such as those of the jamaicamides, bryostatins, curacin A, oocydin, pederin, and psymberin, among others; in biochemically characterized Type I PKS BGCs, HCS-mediated β-branch formation is the common mechanism (Chang et al., 2004; Edwards et al., 2004; Sudek et al., 2007; Fisch et al., 2009; Matilla et al., 2012). Lastly, the hydroxy group on C-10 in the α-position of the acetate subunit, was hypothesized to arise by elongation resulting from hydroxymalonyl-CoA incorporation or by the action of a hydroxylase at a later stage of biosynthesis.

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

The *Ca.* Synoicohabitans palmerolidicus MAG reported a candidate hybrid PKS-NRPS biosynthetic gene cluster that was present in multiple, non-identical copies (Murray et al., 2021). Detailed inspection of one of these clusters (specifically contig 9 which corresponds to *pal* BGC4, the first to be interrogated here) has excellent congruence with the retrobiosynthetic predictions outlined above (**Figure 1**). The results here in which we integrated BGC annotations predicted using AntiSMASH (Blin et al., 2019) with information from protein family homology analysis,

conserved domain searches, active site and motif identification, together support the hypothesis that this 75 kb BGC is putatively responsible for palmerolide A production.

The architecture of the biosynthetic gene cluster reveals core biosynthetic domains followed by 2 ATs, and finally, a series of trans-acting domains (**Figure 2**). The core scaffold is 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 *Ca*. Synoicihabitans palmerolidicus genome. Comparisons of this overall modular architecture with eleven other *trans*-AT systems suggests a significant amount of flexibility in these constructs (**Figure 2**). The mandelalide BGC (Lopera et al. 2017) most closely resembles that of palmerolide in which the core modules are followed in line by AT modules, and trans-acting modules are encoded at the end of the cluster, except that there is only a single AT reported in the case of mandelalide. The proposed BGC for palmerolide A is comprised of 14 core biosynthetic modules and 25 genes in a single operon of 74,655 bases (**Figure 3**). 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 (**Figure 3**).

An unusual starter unit and NRPS domains of *palA.* Bioinformatic analysis of the gene sequence suggests that the initial core biosynthetic domains of *palA* (modules 1 and 2) encode for the requisite acyl carrier proteins (ACP) (**Figure 3**). ACPs are typically responsible for tethering the acyl subunits to a phosopantetheine arm via thioester bond formation. Encoded in module 1 are three ACPs in tandem, which could serve to promote an increase in metabolite production (Gulder et al., 2011). The second in series is an ACP-β containing the conserved domain sequence GXDS (Bertin et al., 2016) which is likely the acceptor of a starter unit containing a β-branch. This is consistent with our proposed starter unit for palmerolide A, 3-methylcrotonic acid. While both *trans*-acting ATs, PalE and PalF, (**Figure 3**) possess the catalytic active site serine which is key for the proper positioning of the selected subunit within the hydrophobic cleft of the active site (Reeves et al., 2001; Helfrich and Piel, 2016), only the first AT, PalE, has a characteristic motif that includes an active site phenylalanine, conferring specificity for malonate selection (Yadav et al., 2003). 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 of this hypothesis, some *trans*-acting ATs have demonstrated affinity for a wider range of substrates than their *cis*-acting counterparts (Dunn et al., 2014; Nivina et al., 2019). 3-methylcrotonyl-CoA is an intermediate of branched-chain amino acid catabolism in leucine degradation; intermediates of this pathway can be diverted to secondary metabolite production (Díaz-Pérez et al., 2016). Interestingly, an upregulation of leucine catabolic pathways can be seen as a mechanism of cold adaption as intermediates, such as methylcrotonyl-CoA can be incorporated into the biosynthetic pathways of branched chain fatty acids (Mahmud et al., 2002; Díaz-Pérez et al., 2016). Integration of branched chain fatty acids in cell walls allows some microorganisms to maintain cell wall fluidity during temperature fluctuations (Klein et al., 1999; Fonseca et al., 2011, 2013), a factor which may be at play in this Antarctic bacterium. The subsequent NRPS module (module 2) contains condensation (C) and adenylation (A) domains as well as a carrier protein. Signature sequence information and NRPSPredictor2 analysis (Röttig et al., 2011; Blin et al., 2019) 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.

In a non-canonical fashion, the carrier proteins flanking the NRPS domains do not appear to be the expected ACP and peptidyl-carrier protein (PCP) for module 1 and 2, respectively. The carrier protein following the KS domain in module 1 was initially annotated as a non-β-branching ACP, however, phylogenetic analysis with the amino acid sequences of carrier proteins from other hybrid PKS-NRPS systems demonstrates that this carrier protein is in the same clade as PCPs (**Supplemental Figure S1**). The carrier protein associated with module 2, which was initially annotated simply as a phosphopantetheine attachment site (Pfam00550.24), is found to be more phylogenetically-related to ACPs within PKS-NRPS systems (**Supplemental Figure S1**). Notably, it possesses the (D/E)xGxDSL motif for phosphopantetheine arm attachment (Keatinge-Clay, 2012) with the exception of an isoleucine rather than leucine in the final position of the motif, which

is a residue common to other ACPs from hybrid PKS-NRPS systems (**Supplemental Figure S2**). Typically, a PCP would follow the domains in NRPS-like modules, however, there are exceptions in the literature. For example, the BGCs for both corallopyronin and oxazolamycin contain ACPs following an A domain (Erol et al., 2010; Zhao et al., 2010). This non-canonical finding could point to the acquisition of these domains over evolutionary time, as the carrier protein for module 1 is encoded in *palA*, the same gene encoding the proteins for both modules 1 and 2, whereas the carrier protein for module 2 is encoded at the beginning of *palB*, a gene which encodes for only PKS domains (**Figure 3**).

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 in which the synthesis proceeds through 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 (**Figure 3**). 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 expected KR domain that would be responsible for the Δ^{22} olefin (**Figure 3**) is not present. The BGCs for bryostatin 1, corallopyronin, and sorangicin also lack an accompanying KR domain to work in concert with an encoded DH. The unaccompanied DH in the bryostatin 1 and corallopyronin systems are deemed inactive; however, an olefin results from the DH in the absence of an accompanying KR in both modules 9 and 11 of the sorangicin BGC (Sudek et al., 2007; Erol et al., 2010; Irschik et al., 2010). 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), 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 substrate critical for macrolactonization of the polyketide is the C-19 hydroxy group, a result of the action of the KS and KR domains encoded in module 5 (**Figure 3**). Interestingly, 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 (Bertin et al., 2016). The DHt sequence does not possess the hotdog fold that is indicative of canonical dehydratases (Cantu et al., 2010), and therefore, may not truly represent a DH. Alternatively, this domain could be responsible for the olefin shifts to the β,γ-positions, as seen in bacillaene and ambruticin biosynthesis (Moldenhauer et al., 2010; Berkhan et al., 2016).

In addition to a standard ACP and KS encoded in module 6, which would lead to a ketone function, an enoyl-CoA hydratase (ECH) is also encoded. Based on our retrobiosynthetic analysis, the ketone at C-17 is the necessary substrate for HCS-catalyzed β-branch formation, resulting in the C-25 methyl group on C-17. We propose that the ECH encoded in module 6 works in concert with the HCS cassette. 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 (**Figure 3**). The HCS cassette can act while the elongating chain is tethered to an ACP module, rather than after cyclization and release (Moldenhauer et al., 2007; Hertweck, 2009). 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 terminal methylene to the observed internal olefin. An HCS cassette formed by the combination of a *trans*-KS and at least one ECH module with an HCS domain is reported in several other bacterial BGCs such as bryostatin 1, calyculin A, jamaicamide, mandelalide, phormidolide, and psymberin (Edwards et al., 2004; Esquenazi et al., 2008; Fisch et al., 2009; Wakimoto et al., 2014; Bertin et al., 2016; Lopera et al., 2017). The domain structure for the HCS cassettes has a remarkably high degree of synteny across these diverse BGCs (Buchholz et al., 2010), however, the presence of an in-line ECH domain in these biosynthetic systems may vary. There is precedence for a similar domain architecture in oocydin, pederin, onnamide, psymberin, phormidolide, and mandelalide, though the presence of the additional ECH domain in-line with the core biosynthetic genes does not necessarily correlate with the formation of an internal versus terminal olefin (Piel et al., 2004; Fisch et al., 2009; Matilla et al., 2012; Bertin et al., 2016; Lopera et al., 2017).

354 There is substantial similarity in the domain structure of module 7, module 10, and module 13,
355 whereby each includes a KS, DH, and KR (**Figure 3**). The olefin that arises from the action of
356 module 7, concomitant with carbon chain elongation, is conjugated with the Δ^{16} olefin adjacent the
357 C-17 β-branch. Modules 10 and 13 have similar enzymatic composition to 7 and are likely
358 responsible for Δ^{8} and Δ^{2} olefins. The combination of KR and DH domains are also found in
359 modules 8 and 12; however, in concert with an as of yet unidentified *trans*-acting enoyl reductase
360 domain (ER), these olefins would be reduced to fully saturated monomeric subunits. There are
361 some examples of *trans*-acting ER domains carrying out this function, including OocU in oocydin,
362 SorN in sorangicin, and MndM in mandelalide (Irschik et al., 2010; Matilla et al., 2012; Lopera et
363 al., 2017), while in other systems, such as corallopyronin and leinamycin, the reductions of the
364 olefins are largely unexplained (Cheng et al., 2003; Erol et al., 2010). The reduction by a *trans*-
365 acting enzyme often occurs while the elongating polyketide is tethered to the megaenzyme, as
366 evidenced by the downstream specificity of the KS module for Claisen-type condensation with
367 subunits containing single or double bonds (Irschik et al., 2010).
368 The genetic architecture for the biosynthesis of two functional groups essential for bioactivity
369 is encoded in module 9 (**Figure 3**). Structure-activity relationship studies demonstrate the
370 importance of the C-10 hydroxyl group and the C-11 carbamate (Nicolaou et al., 2008a). The KR
371 domain predicting the C-11 alcohol function serves as the substrate for the carbamoyl transferase
372 (*palQ*) in a post-translational modification (Haydock et al., 2005; Chen et al., 2009; Mihali et al.,
373 2011). Intriguingly, a domain annotated as a luciferase-like monooxygenase (LLM) initially seems
374 out of place. However, palmerolide A has a hydroxy group at C-10, which is the α-position of the
375 acetate subunit inserted in module 9. LLMs associated with BGCs may not serve as true
376 luciferases, but, instead, demonstrate oxidizing effects on polyketides and peptides without
377 evidence of corresponding bioluminescence (El-Sayed et al., 2001; Maier et al., 2015). For
378 example, there is an overrepresentation of LLMs in *Candidatus* Entotheonella BGCs without known
379 bioluminescence (Lackner et al., 2017). As demonstrated through individual inactivation of the LLM
380 in the BGC of mensacarin, a Type II PKS system, Msn02, Msn04, and Msn08 have key activity as
381 epoxidases and hydroxlases (Maier et al., 2015). There are several examples of LLMs in modular
382 Type I PKS systems. OnnC from onnamide and NazB from nazumamide are two LLMs in
383 *Candidatus* Endotheonella that are proposed to serve biosynthetically as hydroxylases (Lackner et
384 al., 2017). In calyculin and mandelalide, the CalD and MndB LLMs catalyze chain shortening
385 reactions through α-hydroxylation and Baeyer-Villiger-type oxidation reactions (Wakimoto et al.,
386 2014; Lopera et al., 2017). Phormidolide has a LLM that adds a hydroxy group, which is
387 hypothesized to attack an olefin through a Michael-type addition for cyclization with enzymatic
388 assistance from a pyran synthase (Bertin et al., 2016). The hydroxylation that is key in cyclization
389 of oocydin A is likely installed by OocK or OocM, flavin-dependent monooxygenases that are
390 contiguous to the PKS genes and are thought to act while the substrate is bound to a portion of the
391 PKS megaenzyme (Matilla et al., 2012). It is this hydroxylase activity that we propose for the LLM
392 in module 9. Since the producing bacteria is yet to be cultured, it is not established whether this
393 LLM may also serve a role in bioluminescence and/or quorum sensing. Further evidence for the
394 role of the LLM is provided through alignment against other LLM's. The *pal* BGC contains both
395 LuxR family transcriptional regulator as well as the DNA-binding response regulator. In addition to
396 the annotation within Pfam00296, which includes the bacterial LLMs, the sequence, of which, is
397 homologous with the multiple sequence alignments and the hidden Markov models of the TIGR
398 subfamily 04020 of the conserved protein domain family cl19096 noted for natural product
399 biosynthesis LLMs (Lackner et al., 2017). The subfamily occurs in both NRPS and PKS systems
400 as well as small proteins with binding of either flavin mononucleotide or coenzyme F420. Alignment
401 of the LLMs from multiple PKS systems, including palmerolide A, shows homology with model
402 sequences from the TIGR subfamily 04020 (**Supplemental Figure S3**).
403 The addition of C-5 and C-6 and the reduction of the β-carbonyl to form the C-7 hydroxy group
404 of palmerolide A, is due to module 11, which possesses a KR domain in addition to elongating KS
405 (**Figure 3**). In the structure of palmerolide A this is followed by the fully reduced subunit from module
406 12 as discussed above. The final elongation results from module 13, which includes DH and KR

407 domains that contribute to the conjugated ester found as palmerolide A's C-1 through C-3,
408 completing the palmerolide A C_{24} carbon skeleton.
409
410 **Noncanonical termination condensation domain in *palD* for product cyclization and release.**
411 Typically, PKS systems terminate with a thioesterase (TE) domain, leading to release of the
412 polyketide from the megaenzyme (Piel, 2002; Gu et al., 2009; Gehret et al., 2011; Lopera et al.,
413 2017). This canonical domain is not present in the *pal* cluster. Instead, the final module in the *cis*-
414 acting biosynthetic gene cluster includes a truncated condensation domain comprised of 133 amino
415 acid residues, compared to the approximately 450 residues that comprise a standard condensation
416 domain (Stachelhaus et al., 1998) (**Figure 3**). Condensation domains catalyze cyclization through
417 ester formation in free-standing condensation domains that act in *trans* as well as in NRPS systems
418 (Zaleta-Rivera et al., 2006; Lin et al., 2009). In addition, this non-canonical termination domain is
419 not without precedent in hybrid PKS-NRPS and in PKS systems as both basiliskamide and
420 phormidolide include condensation domains for product release (Theodore et al., 2014; Bertin et
421 al., 2016). Though the terminal condensation domain in the *pal* BGC is shortened, it maintains
422 much of the HHXXDDG motif (**Supplemental Figure S3**), most notably the second histidine, which
423 serves as the catalytic histidine in the condensation reaction (Stachelhaus et al., 1998).
424
425 **Stereochemical and structural confirmation based on sequence information.** KR domains are
426 NADPH-dependent enzymes that belong to the short-chain dehydrogenase superfamily, with
427 Rossman-like folds for co-factor binding (Keatinge-Clay and Stroud, 2006; Keatinge-Clay, 2012).
428 Enzymatically, the two KR subtypes, A-Type and B-type, are responsible for stereoselective
429 reduction of β-keto groups and can also determine the stereochemistry of α-substituents. C-type
430 KRs, however, lack reductase activity and often serve as epimerases. A-Type KRs have a key
431 tryptophan residue in the active site, do not possess the LDD amino acid motif, and result in the
432 reduction of β-carbonyls to L-configured hydroxy groups (Keatinge-Clay, 2012). B-Type, which are
433 identified by the presence of an LDD amino acid motif, result in formation of D-configured hydroxy
434 groups (Keatinge-Clay, 2012). The stereochemistry observed in palmerolide A is reflected in the
435 active site sequence information for the D-configured hydroxyl groups from module 5 and module
436 11 (**Figures 1** and **3**). When an enzymatically active DH domain is within the same module, the
437 stereochemistry of the *cis*- versus *trans*-olefin can be predicted, as the combination of an A-Type
438 KR with a DH results in a *cis*-olefin formation and the combination of a B-Type KR with a DH results
439 in *trans*-olefin formation. The *trans*-α,β-olefins arising from module 7 (Δ^{14}), module 10 (Δ^{8}), and
440 module 13 (Δ^{2}) stem from B-Type KRs and active DHs. The other three olefins present in the
441 structure of palmerolide A, as noted above, likely have positional and stereochemical influence
442 during the enzymatic shifts to the β,γ-positions (Δ^{21} and Δ^{23}) or from the ECH domain (module 6).
443 Additional insights into the structural features of the resulting compound were obtained
444 through defining the specificity of KS domains using phylogenetic analysis and the *trans*-AT PKS
445 Polyketide Predictor (*trans*ATor) bioinformatic tool (Helfrich et al., 2019). KS domains catalyze the
446 sequential two-carbon elongation steps through a Claisen-like condensation with a resulting β-keto
447 feature (Khosla et al., 2007). Additional domains within a given module can modify the β-carbonyl
448 or add functionality to the adjacent α- or γ-positions (Keatinge-Clay, 2012). Specificity of KSs,
449 based on the types of modification located on the upstream acetate subunit were determined and
450 found to be mostly consistent with our retrobiosynthetic predictions (**Supplemental Figure S4,**
451 **Supplemental Table S1**). For example, the first KS, KS1 (module 1), is predicted to receive a
452 subunit containing a β-branch. KS3 (module 4) and KS4 (module 5) are predicted to receive an
453 upstream monomeric unit with α-methylation and an olefinic shift, consistent with the structure of
454 palmerolide A and with the enzymatic transformations resulting from module 3 and module 4,
455 respectively. Interestingly, the KS associated with the HCS cassette branches deeply compared to
456 all others upon phylogenetic analysis (**Supplemental Figure S4**). *Trans*ATor also aided in
457 confirming the stereochemical outcomes of the hydroxy groups and olefins, which occur through
458 reduction of the β-carbonyls. The predictions for the D-configured hydroxy groups were consistent
459 with not only the presence of the LDD motif, indicative of type B-type KR as outlined above, but
460 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 *trans*-acting domains and domains between genes responsible for biosynthesis. A glycosyl transferase (*palP*) and lactone oxidase (*palH*) that are often associated with glycosylation of polyketides are encoded in the palmerolide A BGC following the AT domains and preceding the HCS cassette (**Figure 3**). Though glycosylated palmerolides have not been observed, glycosylation as a means of self-resistance in *Streptomyces* has been described and is a possible explanation for these observed domains (Quirós et al., 1998; Wencewicz, 2019). Glycosyl transferases are found in other macrolide- and non-macrolide-producing organisms as a means to inactivate hydroxylated polyketides (Jenkins and Cundliffe, 1991; Gourmelen et al., 1998). Though prokaryotic V-ATPases tend to be more structurally simple than those of eukaryotes, there is homology in the active sites of prokaryotic and eukaryotic V-ATPases making the pro-drug hypothesis for self-resistance reasonable in palmerolide A biosynthesis (Yokoyama and Imamura, 2005). The D-arabinono-1,4-lactone oxidase (*palH*) is a FAD-dependent oxidoreductase that likely works 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 (Murray et al., 2021). This transporter, which has homology to SryD and contains the key nucleotide-binding domain GGNGSGKST, may be responsible for the translocation of the macrolide out of the cell, 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 kingdoms, yet lack assignment of function (Boehm et al., 2009). A domain of unknown function (DUF) is annotated; however, this particular DUF, DUF179, has no superfamily relatives in protein homology searches. Together these genes encoding potential macrolide glycosylation and transport functions may play a role in the bioactivity and export of palmerolide A from the producing organism. Future integrated studies will be needed to decipher the functions of these genes *in situ*.

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

Careful assembly of the Ca. *S. palmerolidicus* MAG revealed the *pal* BGC was present in multiple copies (**Figure 4** and **Supplemental Figure S5-S7**) (Murray et al., 2021), evidenced by their independent anchoring loci within the MAG and supported by a five-fold increase in depth of coverage relative to the rest of the genome. The structural complexity of the multicopy BGCs represents a biosynthetic system that is similar to that found in *Ca*. Didemnitutus mandela, another ascidian-associated verrucomicrobium in the family *Opitutaceae* (Lopera et al., 2017). The five distinct Type-1 PKS BGCs with significant regions of overlap are likely responsible for much of the structural diversity in the family of palmerolides (Diyabalanage et al., 2006; Noguez et al., 2011) (**Figure 4**). Palmerolide A, which is the predominant secondary metabolite isolated from *Synoicum adareanum* (Murray et al., 2020), is hypothesized to arise from the BGC designated as *pal* BGC 4 with additional compounds also arising from this cluster. The other clusters designated as *pal* BGC 1, *pal* BGC 2, *pal* BGC 3, and *pal* BGC 5 and their potential biosynthetic 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 (**Figure 4**) 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) are found in the genome of *Ca.* S. palmerolidicus (Murray et al., 2021), 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 Δ^8 olefin, are either due to a difference of the site of action of the LLM (module 9) or a *trans*-acting hydroxylase. Another member of the compound family, palmerolide C, has structural differences attributable to *trans*-acting enzymes as well. Again, a *trans*-acting hydroxylase or the LLM is proposed to be responsible for hydroxylation on C-8. A hydroxy group on C-9 occurs through reduction of the carbonyl. The carbamate installation occurs on C-11 after *trans*-acting hydroxylation or LLM hydroxylation. In addition, the Δ^8 olefin in palmerolide A is not observed, but rather a Δ^6 olefin.

Additional levels of structural variation are seen at the site of the starter unit, likely due to a level of enzymatic promiscuity of the second AT (PalF). This, combined with differences in the sites of action for the *trans*-acting domains, is likely responsible for the structural differences observed in palmerolide F (**Figure 4**). The terminal olefin on the tail of the macrolide, which perhaps is a product of promiscuity of the selection of the starter unit, the isomeric 3-methyl-3-butenoic acid, is consistent with the aforementioned lack of consensus for malonate selection by the AT. In addition, the KS that receives the starter unit is phylogenetically distinct from the other KS in the *pal* clusters (**Supplemental Figure S4**).

The retrobiosynthetic hypothesis for palmerolide G (**Figure 4**) has much similarity to what is present in *pal* BGC 4; however, the presence of a *cis*-olefin rather than a *trans*-olefin could arise from a difference in the enzymatic activity of module 4. This olefin subsequently undergoes an olefinic shift and, therefore, the stereochemistry is not solely reliant upon the action of the associated KR. Although this difference has not been identified in the BGCs in the samples sequenced, this could be present in other environmental samples that have been batched for processing and compound isolation. Currently, the biosynthetic mechanism is unknown.

The core biosynthetic genes of two palmerolide BGCs (*pal* BGC 1 and *pal* BGC 3) are identical to one another (**Supplemental Figure S5**) and possess an additional elongation module when compared to *pal* BGC 4. Palmerolide D (**Figure 4**) is structurally very similar to palmerolide A with the exception of elongation in the carboxylate tail of the macrolide by an isopropyl group. This could arise from one additional round of starter unit elongation via a KS and methylation. These two identical BGCs are consistent with the additional elongation module found in *pal* BGC1 and 3. The overall architecture and stereochemistry are otherwise maintained. Palmerolide H (**Figure 4**) also likely arises from these two BGCs although it includes the structural differences of both palmerolide B and palmerolide D in which it contains the extended carboxylate tail with a terminal olefin and incorporates hydroxylation on C-8 rather than C-10. Again, there is no genomic evidence that this hydroxylation in the α-position is due to incorporation of hydroxymalonate to explain this but is instead likely due to a *trans*-acting hydroxylase. The carbamate installation occurs on C-7, while sulfonation occurs on C-11 and α-hydroxy placement is on C-8.

The final two *pal* BGCs are shorter with a reduced number of biosynthetic modules found compared to *pal* BGC 4. The gene structure of *pal* BGC 5 (**Supplemental Figure S6**) shows preservation of many of the core biosynthetic genes; however, there are no pre-NRP PKS modules noted in the BGC. The HCS cassette, glycosyl transferase, and CT are all present downstream. The predicted product of this cluster does not correspond with a known palmerolide, though post-translational hydrolysis of the C-24 amide may result in a structure similar to palmerolide E (**Figure 4**), which maintains much of the structure of palmerolide A; however, it is missing the initial polyketide starter unit and the glycine subunit. The final *pal* BGC in *Ca.* S. palmerolidicus, *pal* BGC 2 (**Supplemental Figure S7**), includes only five elongating modules, which would result in a 10-carbon structure that has not been isolated. Interestingly, despite the shortened BGC, the HCS cassette, glycosyl transferase, and CT are all present downstream, and the sequence itself aligns perfectly with few SNPs to the other BGCs (Murray et al., 2021). There would only be a single hydroxy group serving as a substrate for the CT, glycosyl transferase, and sulfatase to act. The 2-carbon site of action for the β-branch introduced in the palmerolide A structures would not be

present. The structure-based retrobiosynthesis of the eight known palmerolides (A-F) can be hypothesized to arise from differences in the core biosynthetic genes of these non-identical copies of the *pal* BGC, starter unit promiscuity, and differing sites of action in the *trans*-acting enzymes.

Conclusion

The putative *pal* BGC has been identified and represents the first BGC elucidated from an Antarctic organism (Murray et al., 2021). As outlined in this retrobiosynthetic strategy, the *pal* BGC represents a *trans*-AT Type I PKS/NRPS hybrid system with compelling alignment to the predicted biosynthetic steps for palmerolide A. The *pal* BGC is proposed to begin with PKS modules resulting in the incorporation of an isovaleric acid derivative, 3-methylcrotonic acid, as a starter unit, followed by incorporation of a glycine residue with NRPS-type modules. Thereafter, eleven rounds of progressive polyketide elongation likely occur and leading to varying degrees of oxidation introduced with each module. There are several interesting non-canonical domains encoded within the BGC, such as an HCS, CT, LLM, and a truncated condensation termination domain. Additionally, a glycosylation domain may be responsible for reversible, pro-drug formation to produce self-resistance to the V-ATPase activity of palmerolide A. There are several additional domains, the function of which have yet to be determined.

A combination of modular alterations, starter unit differences, and activity of *trans*-acting enzymes contributes to Nature's production of a suite of palmerolide analogues. There are a total of five distinct *pal* BGCs in the MAG of *Ca.* S. palmerolidicus, yielding the known eight palmerolides, with genetic differences that explain some of the structural variety seen within this family of compounds. These include differences in modules that comprise the core biosynthetic genes. Additionally, it is proposed that some of the architectural diversity of palmerolides arises from different sites of action of the *trans*-acting, or non-colinear, modules. Starter unit promiscuity is another potential source of the structural differences observed in the compounds. Analysis of the *pal* BGC not only provides insight into the architecture of this Type I PKS/NRPS hybrid BGC with unique features, but also lays the foundational groundwork for drug development studies of palmerolide A via heterologous expression.

Data Deposition

Data referenced here was deposited to the NCBI under BioProject Accession Number PRJNA662631 (44). The GenBank Accession Number for the MAG of *Ca.* Synoicohabitans palmerolidicus is JAGGDC000000000 (44). 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).

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

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.

Funding

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.).

Acknowledgments

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

Berkhan, G., Merten, C., Holec, C., and Hahn, F. (2016). The Interplay between a Multifunctional Dehydratase Domain and a C-Methyltransferase Effects Olefin Shift in Ambruticin Biosynthesis. *Angew. Chemie - Int. Ed.* 55, 13589–13592. doi:10.1002/anie.201607827.

Bertin, M. J., Vulpanovici, A., Monroe, E. A., Korobeynikov, A., Sherman, D. H., Gerwick, L., et al. (2016). The Phormidolide Biosynthetic Gene Cluster: A trans -AT PKS Pathway Encoding a Toxic Macrocyclic Polyketide. *ChemBioChem* 17, 164–173. doi:10.1002/cbic.201500467.

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., et al. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* 47, W81–W87. doi:10.1093/nar/gkz310.

Boehm, M., Nield, J., Zhang, P., Aro, E. M., Komenda, J., and Nixon, P. J. (2009). Structural and mutational analysis of band 7 proteins in the cyanobacterium Synechocystis sp. strain PCC 6803. *J. Bacteriol.* 191, 6425–6435. doi:10.1128/JB.00644-09.

Bordeleau, M. E., Matthews, J., Wojnar, J. M., Lindqvist, L., Novac, O., Jankowsky, E., et al. (2005). Stimulation of mammalian translation initiation factor eIF4A activity by a small molecule inhibitor of eukaryotic translation. *Proc. Natl. Acad. Sci. U. S. A.* 102, 10460–10465. doi:10.1073/pnas.0504249102.

Bowman, E. J., Gustafson, K. R., Bowman, B. J., and Boyd, M. R. (2003). Identification of a New Chondropsin Class of Antitumor Compound that Selectively Inhibits V-ATPases. *J. Biol. Chem.* 278, 44147–44152. doi:10.1074/jbc.M306595200.

Buchholz, T. J., Rath, C. M., Lopanik, N. B., Gardner, N. P., Håkansson, K., and Sherman, D. H. (2010). Polyketide β-Branching in Bryostatin Biosynthesis: Identification of Surrogate Acetyl-ACP Donors for BryR, an HMG-ACP Synthase. *Chem. Biol.* 17, 1092–1100. doi:10.1016/j.chembiol.2010.08.008.

Cantu, D. C., Chen, Y., and Reilly, P. J. (2010). Thioesterases: A new perspective based on their primary and tertiary structures. *Protein Sci.* 19, 1281–1295. doi:10.1002/pro.417.

Carroll, A. R., Copp, B. R., Davis, R. A., Keyzers, R. A., and Prinsep, M. R. (2019). Marine natural products. *Nat. Prod. Rep.* 36, 122–173. doi:10.1039/c8np00092a.

Chang, Z., Sitachitta, N., Rossi, J. V., Roberts, M. A., Flatt, P. M., Jia, J., et al. (2004). 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. doi:10.1021/np0499261.

Chen, W., Huang, T., He, X., Meng, Q., You, D., Bai, L., et al. (2009). Characterization of the Polyoxin Biosynthetic Gene Cluster from Streptomyces cacaoi and Engineered Production

674 of Polyoxin H. *J. Biol. Chem.* 284, 10627–10638. doi:10.1074/jbc.M807534200.
675 Cheng, Y. Q., Tang, G. L., and Shen, B. (2003). Type I polyketide synthase requiring a discrete
676 acyltransferase for polyketide biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* 100, 3149–3154.
677 doi:10.1073/pnas.0537286100.
678 Clarke, A., and Crame, J. A. (2010). Evolutionary dynamics at high latitudes: Speciation and
679 extinction in polar marine faunas. *Philos. Trans. R. Soc. B Biol. Sci.* 365, 3655–3666.
680 doi:10.1098/rstb.2010.0270.
681 Dejong, C. A., Chen, G. M., Li, H., Johnston, C. W., Edwards, M. R., Rees, P. N., et al. (2016).
682 Polyketide and nonribosomal peptide retro-biosynthesis and global gene cluster matching.
683 *Nat. Chem. Biol.* 12, 1007–1014. doi:10.1038/nchembio.2188.
684 Díaz-Pérez, A. L., Díaz-Pérez, C., and Campos-García, J. (2016). Bacterial l-leucine catabolism
685 as a source of secondary metabolites. *Rev. Environ. Sci. Bio/Technology* 15, 1–29.
686 doi:10.1007/s11157-015-9385-3.
687 Diyabalanage, T., Amsler, C. D., McClintock, J. B., and Baker, B. J. (2006). Palmerolide A, a
688 cytotoxic macrolide from the antarctic tunicate Synoicum adareanum. *J. Am. Chem. Soc.*
689 128, 5630–5631. doi:10.1021/ja0588508.
690 Dunn, B. J., Watts, K. R., Robbins, T., Cane, D. E., and Khosla, C. (2014). Comparative analysis
691 of the substrate specificity of trans - Versus cis- acyltransferases of assembly line polyketide
692 synthases. *Biochemistry* 53, 3796–3806. doi:10.1021/bi5004316.
693 Edwards, D. J., Marquez, B. L., Nogle, L. M., McPhail, K., Goeger, D. E., Roberts, M. A., et al.
694 (2004). Structure and Biosynthesis of the Jamaicamides, New Mixed Polyketide-Peptide
695 Neurotoxins from the Marine Cyanobacterium Lyngbya majuscula. *Chem. Biol.* 11, 817–
696 833. doi:10.1016/j.chembiol .2004.03.030.
697 El-Sayed, A. K., Hothersall, J., and Thomas, C. M. (2001). Quorum-sensing-dependent regulation
698 of biosynthesis of the polyketide antiobiotic mupirocin in Pseudomonas fluorescens NCIMB
699 10586. *Microbiology* 147, 2127–2139. doi:10.1099/00221287-147-8-2127.
700 Erol, Ö., Schäberle, T. F., Schmitz, A., Rachid, S., Gurgui, C., El Omari, M., et al. (2010).
701 Biosynthesis of the myxobacterial antibiotic corallopyronin A. *ChemBioChem* 11, 1253–
702 1265. doi:10.1002/cbic.201000085.
703 Esquenazi, E., Coates, C., Simmons, L., Gonzalez, D., Gerwick, W. H., and Dorrestein, P. C.
704 (2008). Visualizing the spatial distribution of secondary metabolites produced by marine
705 cyanobacteria and sponges via MALDI-TOF imaging. *Mol. Biosyst.* 4, 562–570.
706 doi:10.1039/b720018h.
707 Fisch, K. M., Gurgui, C., Heycke, N., Van Der Sar, S. A., Anderson, S. A., Webb, V. L., et al.
708 (2009). Polyketide assembly lines of uncultivated sponge symbionts from structure-based
709 gene targeting. *Nat. Chem. Biol.* 5, 494–501. doi:10.1038/nchembio.176.
710 Fonseca, P., Moreno, R., and Rojo, F. (2011). Growth of Pseudomonas putida at low
711 temperature: global transcriptomic and proteomic analyses. *Environ. Microbiol. Rep.* 3, 329–
712 339. doi:10.1111/j.1758-2229.2010.00229.x.
713 Fonseca, P., Moreno, R., and Rojo, F. (2013). Pseudomonas putida growing at low temperature
714 shows increased levels of CrcZ and CrcY sRNAs, leading to reduced Crc-dependent
715 catabolite repression. *Environ. Microbiol.* 15, 24–35. doi:10.1111/j.1462-2920.2012.02708.x.
716 Gehret, J. J., Gu, L., Gerwick, W. H., Wipf, P., Sherman, D. H., and Smith, J. L. (2011). Terminal
717 alkene formation by the thioesterase of curacin A biosynthesis: Structure of a
718 decarboxylating thioesterase. *J. Biol. Chem.* 286, 14445–14454.
719 doi:10.1074/jbc.M110.214635.
720 Gourmelen, A., Blondelet-Rouault, M. H., and Pernodet, J. L. (1998). Characterization of a
721 glycosyl transferase inactivating macrolides, encoded by gimA from Streptomyces
722 ambofaciens. *Antimicrob. Agents Chemother.* 42, 2612–2619. doi:10.1128/aac.42.10.2612.
723 Greunke, C., Duell, E. R., D'Agostino, P. M., Glöckle, A., Lamm, K., and Gulder, T. A. M. (2018).
724 Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic potential. *Metab.*
725 *Eng.* 47, 334–345. doi:10.1016/j.ymben.2018.03.010.
726 Gu, L., Wang, B., Kulkarni, A., Geders, T. W., Grindberg, R. V., Gerwick, L., et al. (2009).
727 Metamorphic enzyme assembly in polyketide diversification. *Nature* 459, 731–735.

doi:10.1038/nature07870.
Gulder, T. A. M., Freeman, M. F., and Piel, J. (2011). “The Catalytic Diversity of Multimodular Polyketide Synthases: Natural Product Biosynthesis Beyond Textbook Assembly Rules,” in doi:10.1007/128_2010_113.
Haydock, S. F., Aparicio, J. F., Molnár, I., Schwecke, T., Khaw, L. E., König, A., et al. (1995). Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases. *FEBS Lett.* 374, 246–248. doi:10.1016/0014-5793(95)01119-Y.
Haydock, S. F., Appleyard, A. N., Mironenko, T., Lester, J., Scott, N., and Leadlay, P. F. (2005). Organization of the biosynthetic gene cluster for the macrolide concanamycin A in Streptomyces neyagawaensis ATCC 27449. *Microbiology* 151, 3161–3169. doi:10.1099/mic.0.28194-0.
Helfrich, E. J. N., and Piel, J. (2016). Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* 33, 231–316. doi:10.1039/c5np00125k.
Helfrich, E. J. N., Ueoka, R., Dolev, A., Rust, M., Meoded, R. A., Bhushan, A., et al. (2019). Automated structure prediction of trans-acyltransferase polyketide synthase products. *Nat. Chem. Biol.* 15, 813–821. doi:10.1038/s41589-019-0313-7.
Hertweck, C. (2009). The biosynthetic logic of polyketide diversity. *Angew. Chemie - Int. Ed.* 48, 4688–4716. doi:10.1002/anie.200806121.
Irschik, H., Kopp, M., Weissman, K. J., Buntin, K., Piel, J., and Müller, R. (2010). Analysis of the sorangicin gene cluster reinforces the utility of a combined phylogenetic/retrobiosynthetic analysis for deciphering natural product assembly by trans-AT PKS. *ChemBioChem* 11, 1840–1849. doi:10.1002/cbic.201000313.
Jenke-Kodama, H., Sandmann, A., Müller, R., and Dittmann, E. (2005). Evolutionary implications of bacterial polyketide synthases. *Mol. Biol. Evol.* 22, 2027–2039. doi:10.1093/molbev/msi193.
Jenkins, G., and Cundliffe, E. (1991). Cloning and characterization of two genes from Streptomyces lividans that confer inducible resistance to lincomycin and macrolide antibiotics. *Gene* 108, 55–62. doi:10.1016/0378-1119(91)90487-V.
Jiang, X., Liu, B., Lebreton, S., and De Brabander, J. K. (2007). Total synthesis and structure revision of the marine metabolite palmerolide A. *J. Am. Chem. Soc.* 129, 6386–6387. doi:10.1021/ja0715142.
Kaliappan, K. P., and Gowrisankar, P. (2007). Synthetic studies on a marine natural product, palmerolide A: Synthesis of C1-C9 and C15-C21 fragments. *Synlett*, 1537–1540. doi:10.1055/s-2007-982539.
Kautsar, S. A., Blin, K., Shaw, S., Navarro-Muñoz, J. C., Terlouw, B. R., Van Der Hooft, J. J. J., et al. (2020). MIBiG 2.0: A repository for biosynthetic gene clusters of known function. *Nucleic Acids Res.* 48, D454–D458. doi:10.1093/nar/gkz882.
Kazami, S., Muroi, M., Kawatani, M., Kubota, T., Usui, T., Kobayashi, J., et al. (2006). Iejimalides show anti-osteoclast activity via V-ATPase inhibition. *Biosci. Biotechnol. Biochem.* 70, 1364–1370. doi:10.1271/bbb.50644.
Keatinge-Clay, A. T. (2012). The structures of type I polyketide synthases. *Nat. Prod. Rep.* 29, 1050. doi:10.1039/c2np20019h.
Keatinge-Clay, A. T., and Stroud, R. M. (2006). The Structure of a Ketoreductase Determines the Organization of the β-Carbon Processing Enzymes of Modular Polyketide Synthases. *Structure* 14, 737–748. doi:10.1016/j.str.2006.01.009.
Khosla, C., Tang, Y., Chen, A. Y., Schnarr, N. A., and Cane, D. E. (2007). Structure and Mechanism of the 6-Deoxyerythronolide B Synthase. *Annu. Rev. Biochem.* 76, 195–221. doi:10.1146/annurev.biochem.76.053105.093515.
Kim, E. J., Lee, J. H., Choi, H., Pereira, A. R., Ban, Y. H., Yoo, Y. J., et al. (2012). Heterologous Production of 4-O-Demethylbarbamide, a Marine Cyanobacterial Natural Product. *Org. Lett.* 14, 5824–5827. doi:10.1021/ol302575h.
Klein, W., Weber, M. H. W., and Marahiel, M. A. (1999). Cold shock response of Bacillus subtilis: Isoleucine-dependent switch in the fatty acid branching pattern for membrane adaptation to

782 low temperatures. *J. Bacteriol.* 181, 5341–5349. doi:10.1128/jb.181.17.5341-5349.1999.
783 Koskinen, A. M. P., and Karisalmi, K. (2005). Polyketide stereotetrads in natural products. *Chem.*
784 *Soc. Rev.* 34, 677–690. doi:10.1039/b417466f.
785 Lackner, G., Peters, E. E., Helfrich, E. J. N., and Piel, J. (2017). Insights into the lifestyle of
786 uncultured bacterial natural product factories associated with marine sponges. *Proc. Natl.*
787 *Acad. Sci.* 114, E347–E356. doi:10.1073/pnas.1616234114.
788 Lawrence, J. G., Ochman, H., and Ragan, M. A. (2002). Reconciling the many faces of lateral
789 gene transfer. *Trends Microbiol.* 10, 1–4. doi:10.1016/S0966-842X(01)02282-X.
790 Lebar, M. D., and Baker, B. J. (2010). Synthesis of the C3-14 fragment of palmerolide A using a
791 chiral pool based strategy. *Tetrahedron* 66, 1557–1562. doi:10.1016/j.tet.2009.12.007.
792 Lin, S., Van Lanen, S. G., and Shen, B. (2009). A free-standing condensation enzyme catalyzing
793 ester bond formation in C-1027 biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* 106, 4183–
794 4188. doi:10.1073/pnas.0808880106.
795 Lisboa, M. P., Jones, D. M., and Dudley, G. B. (2013). Formal synthesis of palmerolide A,
796 featuring alkynogenic fragmentation and syn-selective vinylogous aldol chemistry. *Org. Lett.*
797 15, 886–889. doi:10.1021/ol400014e.
798 Lopera, J., Miller, I. J., McPhail, K. L., and Kwan, J. C. (2017). Increased Biosynthetic Gene
799 Dosage in a Genome-Reduced Defensive Bacterial Symbiont. *mSystems* 2.
800 doi:10.1128/mSystems.00096-17.
801 Mahmud, T., Bode, H. B., Silakowski, B., Kroppenstedt, R. M., Xu, M., Nordhoff, S., et al. (2002).
802 A novel biosynthetic pathway providing precursors for fatty acid biosynthesis and secondary
803 metabolite formation in myxobacteria. *J. Biol. Chem.* 277, 32768–32774.
804 doi:10.1074/jbc.M205222200.
805 Maier, S., Heitzler, T., Asmus, K., Brötz, E., Hardter, U., Hesselbach, K., et al. (2015). Functional
806 characterization of different ORFs including luciferase-like monooxygenase genes from the
807 mensacarcin gene cluster. *ChemBioChem* 16, 1175–1182. doi:10.1002/cbic.201500048.
808 Matilla, M. A., Stöckmann, H., Leeper, F. J., and Salmond, G. P. C. (2012). Bacterial biosynthetic
809 gene clusters encoding the anti-cancer haterumalide class of molecules: Biogenesis of the
810 broad spectrum antifungal and anti-oomycete compound, oocydin A. *J. Biol. Chem.* 287,
811 39125–39138. doi:10.1074/jbc.M112.401026.
812 Mattheus, W., Gao, L. J., Herdewijn, P., Landuyt, B., Verhaegen, J., Masschelein, J., et al.
813 (2010). Isolation and Purification of a New Kalimantacin/Batumin-Related Polyketide
814 Antibiotic and Elucidation of Its Biosynthesis Gene Cluster. *Chem. Biol.* 17, 149–159.
815 doi:10.1016/j.chembiol.2010.01.014.
816 McClintock, J. B., Amsler, C. D., Baker, B. J., and Van Soest, R. W. M. (2005). Ecology of
817 antarctic marine sponges: An overview. *Integr. Comp. Biol.* 45, 359–368.
818 doi:10.1093/icb/45.2.359.
819 Mihali, T. K., Carmichael, W. W., and Neilan, B. A. (2011). A putative gene cluster from a
820 Lyngbya wollei bloom that encodes paralytic shellfish toxin biosynthesis. *PLoS One* 6.
821 doi:10.1371/journal.pone.0014657.
822 Moldenhauer, J., Chen, X. H., Borriss, R., and Piel, J. (2007). Biosynthesis of the antibiotic
823 bacillaene, the product of a giant polyketide synthase complex of the trans-AT family.
824 *Angew. Chemie - Int. Ed.* 46, 8195–8197. doi:10.1002/anie.200703386.
825 Moldenhauer, J., Götz, D. C. G., Albert, C. R., Bischof, S. K., Schneider, K., Süssmuth, R. D., et
826 al. (2010). The Final Steps of Bacillaene Biosynthesis in Bacillus amyloliquefaciens FZB42:
827 Direct Evidence for β,γ Dehydration by a trans-Acyltransferase Polyketide Synthase.
828 *Angew. Chemie* 122, 1507–1509. doi:10.1002/ange.200905468.
829 Murray, A. E., Avalon, N. E., Bishop, L., Davenport, K. W., Delage, E., Dichosa, A. E. K., et al.
830 (2020). Uncovering the Core Microbiome and Distribution of Palmerolide in Synoicum
831 adareanum Across the Anvers Island Archipelago, Antarctica. *Mar. Drugs* 18, 298.
832 doi:10.3390/md18060298.
833 Murray, A. E., Lo, C.-C., Daligault, H., Avalon, N. E., Read, R., Davenport, K. W., et al. (2021).
834 Discovery of an Antarctic ascidian-associated uncultivated verrucomicrobium that encodes
835 the biosynthetic capability for antimelanoma palmerolides. (Accepted at mSphere; preprint

available on BioRxiv: doi:https://doi.org/10.1101/2021.05.05.442870).
Napolitano, J., Daranas, A., Norte, M., and Fernandez, J. (2012). Marine Macrolides, a Promising Source of Antitumor Compounds. *Anticancer. Agents Med. Chem.* 9, 122–137. doi:10.2174/187152009787313800.
Nguyen, T. A., Ishida, K., Jenke-Kodama, H., Dittmann, E., Gurgui, C., Hochmuth, T., et al. (2008). Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat. Biotechnol.* 26, 225–233. doi:10.1038/nbt1379.
Nicolaou, K. C., Leung, G. Y. C., Dethe, D. H., Guduru, R., Sun, Y. P., Chek, S. L., et al. (2008a). Chemical synthesis and biological evaluation of palmerolide A analogues. *J. Am. Chem. Soc.* 130, 10019–10023. doi:10.1021/ja802803e.
Nicolaou, K. C., Sun, Y. P., Guduru, R., Banerji, B., and Chen, D. Y. K. (2008b). Total synthesis of the originally proposed and revised structures of palmerolide A and isomers thereof. *J. Am. Chem. Soc.* 130, 3633–3644. doi:10.1021/ja710485n.
Nishimura, S., Matsunaga, S., Yoshida, S., Nakao, Y., Hirota, H., and Fusetani, N. (2005). Structure-activity relationship study on 13-deoxytedanolide, a highly antitumor macrolide from the marine sponge Mycale adhaerens. *Bioorganic Med. Chem.* 13, 455–462. doi:10.1016/j.bmc.2004.10.014.
Nivina, A., Yuet, K. P., Hsu, J., and Khosla, C. (2019). Evolution and Diversity of Assembly-Line Polyketide Synthases. *Chem. Rev.* 119, 12524–12547. doi:10.1021/acs.chemrev.9b00525.
Noguez, J. H., Diyabalanage, T. K. K., Miyata, Y., Xie, X. S., Valeriote, F. A., Amsler, C. D., et al. (2011). Palmerolide macrolides from the Antarctic tunicate Synoicum adareanum. *Bioorganic Med. Chem.* 19, 6608–6614. doi:10.1016/j.bmc.2011.06.004.
Paull, K. D., Hamel, E., and Malspeis, L. (1995). Prediction of Biochemical Mechanism of Action from the In Vitro Antitumor Screen of the National Cancer Institute. *Cancer Chemother. Agents*, 9–45. Available at: https://dtp.cancer.gov/databases_tools/docs/compare/compare.htm [Accessed October 10, 2018].
Pawar, A. B., and Prasad, K. R. (2012). Formal total synthesis of palmerolide A. *Chem. - A Eur. J.* 18, 15202–15206. doi:10.1002/chem.201202324.
Penner, M., Rauniyar, V., Kaspar, L. T., and Hall, D. G. (2009). Catalytic asymmetric synthesis of palmerolide a via organoboron methodology. *J. Am. Chem. Soc.* 131, 14216–14217. doi:10.1021/ja906429c.
Piel, J. (2002). A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles. *Proc. Natl. Acad. Sci. U. S. A.* 99, 14002–14007. doi:10.1073/pnas.222481399.
Piel, J., Hui, D., Wen, G., Butzke, D., Platzer, M., Fusetani, N., et al. (2004). Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge Theonella swinhoei. *Proc. Natl. Acad. Sci. U. S. A.* 101, 16222–16227. doi:10.1073/pnas.0405976101.
Pujari, S. A., Gowrisankar, P., and Kaliappan, K. P. (2011). A shimizu non-aldol approach to the formal total synthesis of palmerolide A. *Chem. - An Asian J.* 6, 3137–3151. doi:10.1002/asia.201100429.
Quirós, L. M., Aguirrezabalaga, I., Olano, C., Méndez, C., and Salas, J. A. (1998). Two glycosyltransferases and a glycosidase are involved in oleandomycin modification during its biosynthesis by Streptomyces antibioticus. *Mol. Microbiol.* 28, 1177–1185. doi:10.1046/j.1365-2958.1998.00880.x.
Ravenhall, M., Škunca, N., Lassalle, F., and Dessimoz, C. (2015). Inferring Horizontal Gene Transfer. *PLoS Comput. Biol.* 11, 1–16. doi:10.1371/journal.pcbi.1004095.
Reeves, C. D., Murli, S., Ashley, G. W., Piagentini, M., Hutchinson, C. R., and McDaniel, R. (2001). Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site-specific mutations. *Biochemistry* 40, 15464–15470. doi:10.1021/bi015864r.
Riesenfeld, C. S., Murray, A. E., and Baker, B. J. (2008). Characterization of the microbial community and polyketide biosynthetic potential in the palmerolide-producing tunicate

890 Synoicum adareanum. *J. Nat. Prod.* 71, 1812–1818. doi:10.1021/np800287n.
891 Röttig, M., Medema, M. H., Blin, K., Weber, T., Rausch, C., and Kohlbacher, O. (2011).
892 NRPSpredictor2 - A web server for predicting NRPS adenylation domain specificity. *Nucleic*
893 *Acids Res.* 39, 362–367. doi:10.1093/nar/gkr323.
894 Schmidt, E. W. (2008). Trading molecules and tracking targets in symbiotic interactions. *Nat.*
895 *Chem. Biol.* 4, 466–473. doi:10.1038/nchembio.101.
896 Schmidt, E. W. (2015). The secret to a successful relationship: Lasting chemistry between
897 ascidians and their symbiotic bacteria. *Invertebr. Biol.* 134, 88–102. doi:10.1111/ivb.12071.
898 Schmitt, I., and Lumbsch, H. T. (2009). Ancient horizontal gene transfer from bacteria enhances
899 biosynthetic capabilities of fungi. *PLoS One* 4, 1–8. doi:10.1371/journal.pone.0004437.
900 Shen, B., Cheng, Y. Q., Christenson, S. D., Jiang, H., Ju, J., Kwon, H. J., et al. (2007). Polyketide
901 biosynthesis beyond the type I, II, and III polyketide synthase paradigms: A progress report.
902 *ACS Symp. Ser.* 955, 154–166. doi:10.1021/bk-2007-0955.ch011.
903 Shen, R., Lin, C. T., Bowman, E. J., Bowman, B. J., and Porco, J. A. (2003). Lobatamide C: Total
904 synthesis, stereochemical assignment, preparation of simplified analogues, and V-ATPase
905 inhibition studies. *J. Am. Chem. Soc.* 125, 7889–7901. doi:10.1021/ja0352350.
906 Stachelhaus, T., Mootz, H. D., Bergendahl, V., and Marahiel, M. A. (1998). Peptide Bond
907 Formation in Nonribosomal Peptide Biosynthesis. *J. Biol. Chem.* 273, 22773–22781.
908 doi:10.1074/jbc.273.35.22773.
909 Sudek, S., Lopanik, N. B., Waggoner, L. E., Hildebrand, M., Anderson, C., Liu, H., et al. (2007).
910 Identification of the putative bryostatin polyketide synthase gene cluster from “Candidatus
911 Endobugula sertula”, the uncultivated microbial symbiont of the marine bryozoan Bugula
912 neritina. *J. Nat. Prod.* 70, 67–74. doi:10.1021/np060361d.
913 Tatsuno, S., Arakawa, K., and Kinashi, H. (2007). Analysis of Modular-iterative Mixed
914 Biosynthesis of Lankacidin by Heterologous Expression and Gene Fusion. *J. Antibiot.*
915 *(Tokyo).* 60, 700–708. doi:10.1038/ja.2007.90.
916 Theodore, C. M., Stamps, B. W., King, J. B., Price, L. S. L., Powell, D. R., Stevenson, B. S., et al.
917 (2014). Genomic and metabolomic insights into the natural product biosynthetic diversity of
918 a feral-hog-associated Brevibacillus laterosporus strain. *PLoS One* 9, 3–12.
919 doi:10.1371/journal.pone.0090124.
920 Ueoka, R., Uria, A. R., Reiter, S., Mori, T., Karbaum, P., Peters, E. E., et al. (2015). Metabolic
921 and evolutionary origin of actin-binding polyketides from diverse organisms. *Nat. Chem.*
922 *Biol.* 11, 705–712. doi:10.1038/nchembio.1870.
923 Videau, P., Wells, K. N., Singh, A. J., Gerwick, W. H., and Philmus, B. (2016). Assessment of
924 Anabaena sp. Strain PCC 7120 as a Heterologous Expression Host for Cyanobacterial
925 Natural Products: Production of Lyngbyatoxin A. *ACS Synth. Biol.* 5, 978–988.
926 doi:10.1021/acssynbio.6b00038.
927 Von Schwarzenberg, K., Wiedmann, R. M., Oak, P., Schulz, S., Zischka, H., Wanner, G., et al.
928 (2013). Mode of cell death induction by pharmacological vacuolar H +-ATPase (V-ATPase)
929 inhibition. *J. Biol. Chem.* 288, 1385–1396. doi:10.1074/jbc.M112.412007.
930 Wakimoto, T., Egami, Y., Nakashima, Y., Wakimoto, Y., Mori, T., Awakawa, T., et al. (2014).
931 Calyculin biogenesis from a pyrophosphate protoxin produced by a sponge symbiont. *Nat.*
932 *Chem. Biol.* 10, 648–655. doi:10.1038/nchembio.1573.
933 Wang, B., Guo, F., Huang, C., and Zhao, H. (2020). Unraveling the iterative type I polyketide
934 synthases hidden in Streptomyces. *Proc. Natl. Acad. Sci. U. S. A.* 117, 8449–8454.
935 doi:10.1073/pnas.1917664117.
936 Wencewicz, T. A. (2019). Crossroads of Antibiotic Resistance and Biosynthesis. *J. Mol. Biol.* 431,
937 3370–3399. doi:10.1016/j.jmb.2019.06.033.
938 Wilkinson, B., Foster, G., Rudd, B. A. M., Taylor, N. L., Blackaby, A. P., Sidebottom, P. J., et al.
939 (2000). Novel octaketide macrolides related to 6-deoxyerythronolide B provide evidence for
940 iterative operation of the erythromycin polyketide synthase. *Chem. Biol.* 7, 111–117.
941 doi:10.1016/S1074-5521(00)00076-4.
942 Yadav, G., Gokhale, R. S., and Mohanty, D. (2003). Computational approach for prediction of
943 domain organization and substrate specificity of modular polyketide synthases. *J. Mol. Biol.*

328, 335–363. doi:10.1016/S0022-2836(03)00232-8.
Yokoyama, K., and Imamura, H. (2005). Rotation, structure, and classification of prokaryotic V-ATPase. *J. Bioenerg. Biomembr.* 37, 405–410. doi:10.1007/s10863-005-9480-1.
Young, R. M., Von Salm, J. L., Amsler, M. O., Lopez-Bautista, J., Amsler, C. D., McClintock, J. B., et al. (2013). Site-specific variability in the chemical diversity of the Antarctic red alga Plocamium cartilagineum. *Mar. Drugs* 11, 2126–2139. doi:10.3390/md11062126.
Zaleta-Rivera, K., Xu, C., Yu, F., Butchko, R. A. E., Proctor, R. H., Hidalgo-Lara, M. E., et al. (2006). A bidomain nonribosomal peptide synthetase encoded by FUM14 catalyzes the formation of tricarballylic esters in the biosynthesis of fumonisins. *Biochemistry* 45, 2561–2569. doi:10.1021/bi052085s.
Zhao, C., Coughlin, J. M., Ju, J., Zhu, D., Wendt-Pienkowski, E., Zhou, X., et al. (2010). Oxazolomycin biosynthesis in Streptomyces albus JA3453 featuring an “acyltransferase-less” type I polyketide synthase that incorporates two distinct extender units. *J. Biol. Chem.* 285, 20097–20108. doi:10.1074/jbc.M109.090092.

Figures

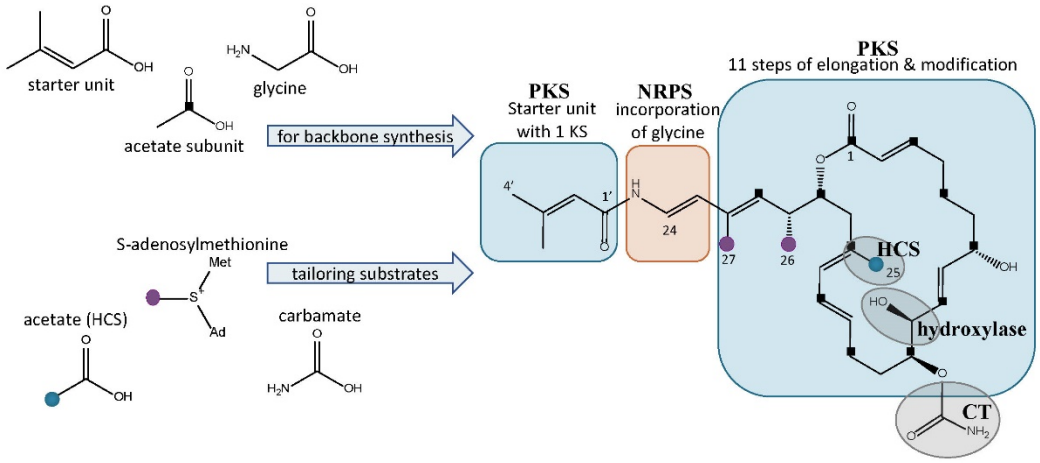
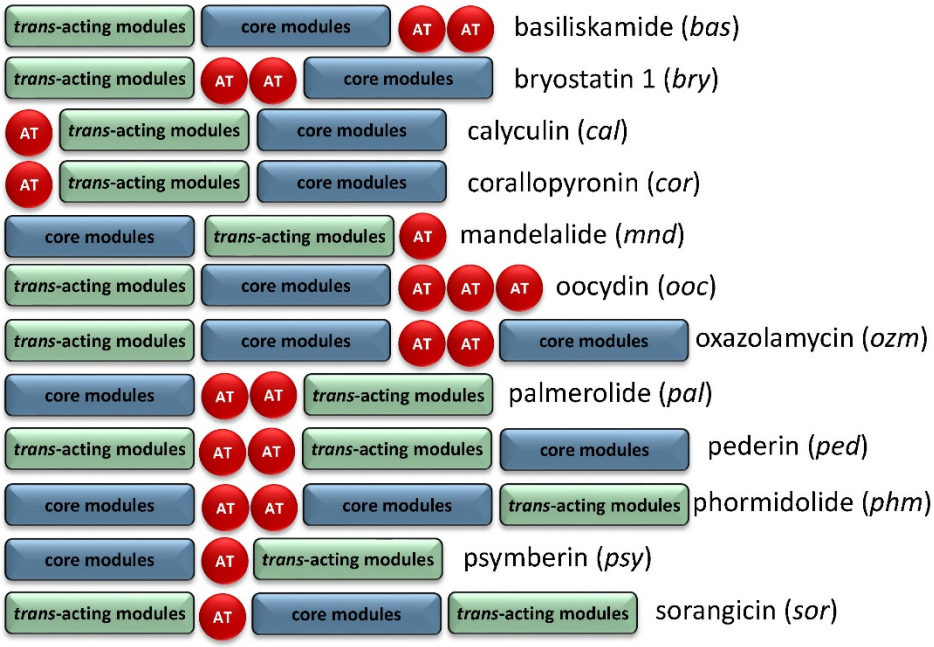
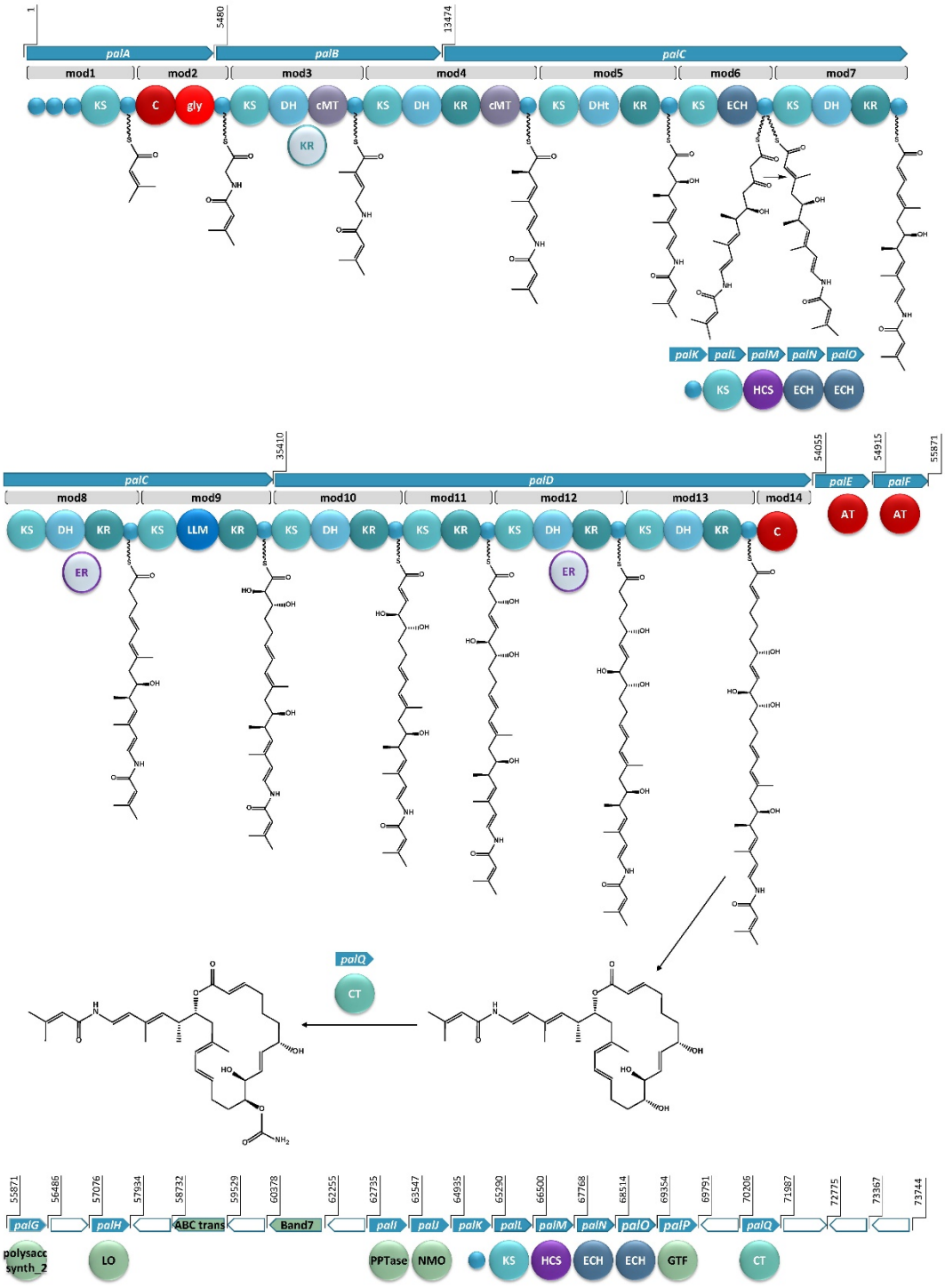


Figure 1. Structure of palmerolide A with notations for the proposed retrobiosynthesis. Backbone synthesis is a result of incorporation of the starter unit, a glycine residue, and acetate subunits (C1 indicated by black squares). Structural features from *trans*-acting tailoring enzymes (indicated by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

970



972 **Figure 2.** Comparison of BGC organization of select *trans*-AT systems. There is significant
973 variability in the order of the core modules, AT modules, and modules which contain *trans*-acting
974 tailoring enzymes. There is also variability in the number of encoded AT modules, though the AT
975 modules are typically encoded on separate, but tandem genes if more than one is present.
976


palA
palB
palC
mod1
mod2
mod3
mod4
mod5
mod6
mod7
KS
C
gly
DH
cMT
KR
DHt
ECH
palK
palL
palM
palN
palO
HCS
palD
palE
palF
mod8
mod9
mod10
mod11
mod12
mod13
mod14
LLM
ER
AT
palQ
CT
palG
palH
palI
palJ
palP
ABC trans
Band7
polysacc synth_2
LO
PPTase
NMO
GTF

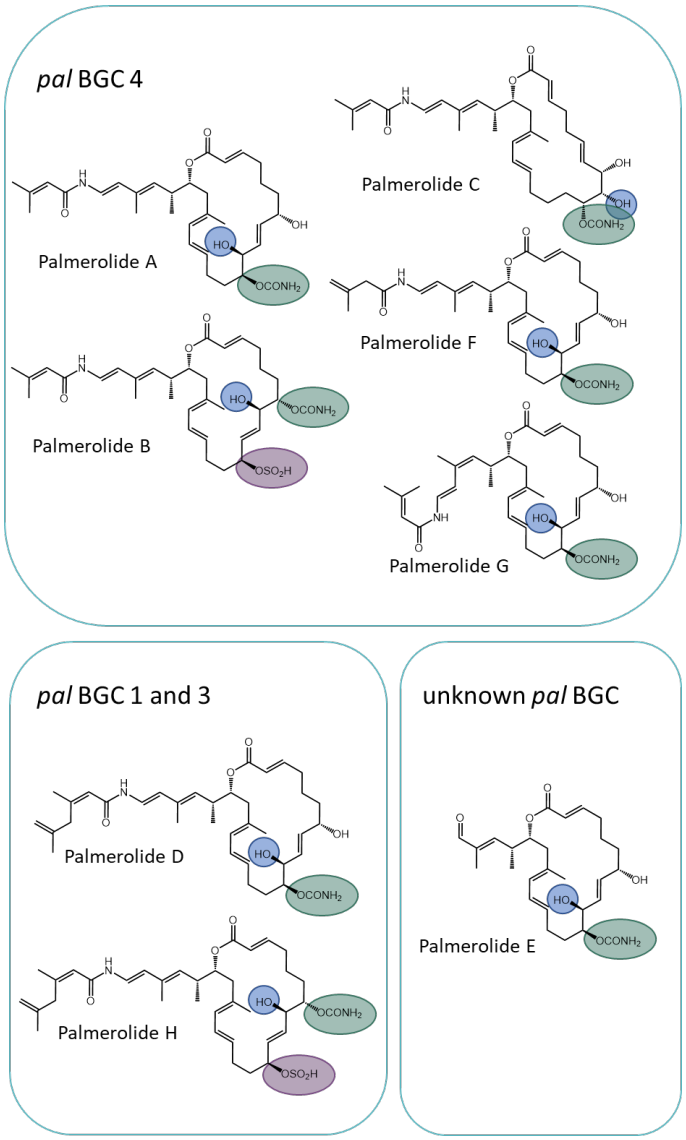
978

979 **Figure 3.** The proposed BGC for palmerolide A, showing the hybrid PKS-NRPS system. KS:
980 ketosynthase domain, C: condensation domain, gly: adenylation domain for glycine incorporation,
981 DH: dehydratase domain, cMT: carbon methyl transferase domain, KR: ketoreductase domain,
982 DHt: dehydratase variant; ECH: enoyl-CoA hydratase, LLM: luciferase-like monooxygenase, AT:
983 acyl transferase; polysacc synt_2: polysaccharide biosynthesis protein, LO: lactone oxidase, ABC
984 trans: ATP-binding cassette transporter, Band7: stomatin-like integral membrane, PPTase:
985 phosphopantetheinyl transferase, NMO: nitronate monooxygenase, HCS: hydroxymethylglutaryl-
986 CoA synthase, GTF: glycosyl transferase ER: enoyl reductase, CT: carbamoyl transferase, small
987 blue circles represent acyl- or peptidyl-carrier proteins. Ppant arms are symbolized by wavy lines.
988 The grey domains indicate domains that would be expected to perform an enzymatic
989 transformation; however, are not encoded in the BGC. Blue arrows indicate biosynthetic genes.
990 Green arrows indicate genes that encode for non-biosynthetic proteins. White arrows reflect
991 hypothetical genes. The BGC is displayed in reverse compliment.
992
993

994 **(A)**

pal BGC 1
KS KS C-gly KS-DH-cMT KS-DH-KR-cMT KS-DHt-KR KS-ECH KS-DH-KR KS-DH-KR KS-LLM-KR KS-DH-KR KS-KR KS-DH-KR KS-DH-KR C
AT AT KS-HCS-ECH-ECH CT
pal BGC 2
KS-DH-KR KS-KR KS-DH-KR KS-DH-KR C
AT AT KS-HCS-ECH-ECH CT
pal BGC 3
KS KS C-gly KS-DH-cMT KS-DH-KR-cMT KS-DHt-KR KS-ECH KS-DH-KR KS-DH-KR KS-LLM-KR KS-DH-KR KS-KR KS-DH-KR KS-DH-KR C
AT AT KS-HCS-ECH-ECH CT
pal BGC 4
KS C-gly KS-DH-cMT KS-DH-KR-cMT KS-DHt-KR KS-ECH KS-DH-KR KS-DH-KR KS-LLM-KR KS-DH-KR KS-KR KS-DH-KR KS-DH-KR C
AT AT KS-HCS-ECH-ECH CT
pal BGC 5
C-gly KS-DH-cMT KS-DH-KR-cMT KS-DHt-KR KS-ECH KS-DH-KR KS-DH-KR KS-LLM-KR KS-DH-KR KS-KR KS-DH-KR KS-DH-KR C
AT AT KS-HCS-ECH-ECH CT

996 **(B)**


pal BGC 4
Palmerolide A
Palmerolide B
Palmerolide C
Palmerolide F
Palmerolide G
pal BGC 1 and 3
Palmerolide D
Palmerolide H
unknown pal BGC
Palmerolide E

998
999 **Figure 4 (A)** Comparison of the modular structure of the 5 *pal* BGCs. **(B)** Family of palmerolides.
1000 Much of the structural diversity can be explained by differences due to starter unit promiscuity,
1001 sites of action for the *trans*-acting tailoring enzymes, and differences in the core modules of the
1002 multiple *pal* BGCs. It is proposed that *pal* BGC 4 is responsible for not only palmerolide A, but
1003 also palmerolide B, palmerolide C, palmerolide F, and palmerolide G. It is interesting to note that
1004 the modular structure of the domains responsible for biosynthesis are equivalent for *pal* BGC 1
1005 and *pal* BGC 3. These two BGCs contain an additional KS domain as compared to *pal* BGC 4
1006 and are likely responsible for the biosynthesis of palmerolide D and palmerolide H.