

**Critical analysis of polycyclic tetramate macrolactam biosynthetic gene cluster
phylogeny and functional diversity**

Christopher P. Harper^{a,*}, Anna Day^{a,*}, Maya Tsingos^a, Edward Ding^{a,b}, Elizabeth Zeng^a,
Spencer D. Stumpf^{a,c}, Yunci Qi^{a,d}, Adam Robinson^{a,e}, Jennifer Greif^a, Joshua A. V.
Blodgett^{a#}

^aDepartment of Biology, Washington University in St Louis. St Louis, MO

Current addresses:

^bColumbia University School of Dental Medicine, New York, NY

^cPfizer Inc, Chesterfield, Missouri

^dUSDA-ARS, New Orleans, LA

^eTulane School of Medicine, New Orleans, LA

*Equal contribution

[#]To whom correspondence should be addressed:

Department of Biology, Washington University in St Louis, 133 Reibstock Hall, 1
Brookings Dr. CB#1137, St Louis MO 63130. Email: jblodgett@wustl.edu Ph: 1 314 935
6233 orcid.org/0000-0002-7080-5870

Running Title: Polycyclic Tetramate Macrolactam BGC Diversity

ABSTRACT

Polycyclic tetramate macrolactams (PTMs or PoTeMs) are bioactive natural products commonly associated with certain actinobacterial and proteobacterial lineages. These molecules have been the subject of numerous structure-activity investigations since the 1970s. New members continue to be pursued in wild and engineered bacterial strains, and advances in PTM biosynthesis suggests their outwardly simplistic biosynthetic gene clusters (BGCs) belie unexpected product complexity. Towards addressing the origins of this complexity and understanding its influence on PTM discovery, we engaged in a combination of bioinformatics to systematically classify PTM BGCs, and PTM- targeted metabolomics to compare the products of select BGC types. By comparing groups of producers and BGC mutants, we exposed knowledge gaps that complicate bioinformatics-driven product predictions. In sum, we provide new insights into the evolution of PTM BGCs while systematically accounting for the PTMs discovered thus far. The combined computational and metabologenomic findings presented here should prove useful for guiding future discovery.

IMPORTANCE: Polycyclic tetramate macrolactam (PTM) pathways are frequently found within the genomes of biotechnologically- important bacteria, including *Streptomyces* and *Lysobacter* spp. Their molecular products are typically bioactive, having substantial agricultural and therapeutic interest. Leveraging bacterial genomics for the discovery of new related molecules is thus desirable, but drawing accurate structural predictions from bioinformatics alone remains challenging. This difficulty stems from a combination of previously underappreciated biosynthetic complexity and remaining knowledge gaps, compounded by a stream of yet- uncharacterized PTM biosynthetic loci gleaned from

recently sequenced bacterial genomes. We engaged in the following study to create a useful framework for cataloging historic PTM clusters, identifying new cluster variations, and tracing evolutionary paths for these molecules. Our data suggests new PTM chemistry remains discoverable in nature. However, our metabolomic and mutational analyses emphasize practical limitations to genomics-based discovery by exposing hidden complexity.

KEYWORDS: polycyclic tetramate macrolactams, phylogeny, biosynthetic gene clusters, comparative metagenomics

INTRODUCTION

Polycyclic tetramate macrolactams (abbreviated PTMs or PoTeMs) are widely studied antibiotic-like bacterial natural products. They were first discovered in the bioactive fractions of microbial and marine life extracts, but their biosyntheses remained uncharacterized until being linked to discrete biosynthetic gene clusters (BGCs) in *Lysobacter enzymogenes* C3 (1) and *Streptomyces* sp. strain SPB78 (2). Consequently, PTMs embody a relatively scarce example of an antibiotic family produced by both Gram-positive and Gram-negative organisms (3, 4). Over time, reports of PTM molecules and their encoding BGCs expanded beyond *Streptomyces* and *Lysobacter* to include several other filamentous actinomycetes (*Kitasatospora*, *Salinispora*, *Micromonospora*, *Saccharopolyspora*, *Saccharothrix*, *Umezawaea*, *Nocardiopsis*, *Actinokineospora*, and *Actinoalloteichus*), gamma-proteobacteria (*Pseudoalteromonas*, *Colwellia*, *Gynuella*, and *Saccharophagus*), as well as the firmicute *Pelosinus fermentans* (1, 2, 5–14).

Regardless of producer phylogeny, all characterized PTM biosyntheses are initiated via an unusual hybrid-iterative polyketide synthase/ non-ribosomal peptide synthetase (iPKS-NRPS, encoded by *ftdB*, *ikaA*, and orthologs) that produces the polyene-tetramate precursor lysobacterene A (or related isomers) (15, 16). However, individual PTM biosynthetic pathways significantly diverge from this point, generally via combinations of oxidative decorations and carbocyclic ring arrangements (Fig 1A). There are now >85 known PTM congeners, and exploring PTM diversity has led to a growing body of valuable structure-activity relationship knowledge (17–19, 5, 20–22, 10, 23–26, 11). Testament to their structural diversity and their presumed interaction with a spectrum of biological targets, PTMs display a range of bioactivities. Examples include anti-inflammatory (27), antifungal (2), antibacterial (28), and predicted anticancer (29) activities, making the family compelling for future pharmaceutical and crop protective investigation. PTM BGCs are often targeted for synthetic refactoring to create new chemical matter or to probe their transcriptional regulation towards increasing production (10, 18, 22, 24, 30–32).

PTM BGCs typically feature relatively low gene content (3-7 genes), organized in operon-like arrangements clustered around their *ftdB* homologs. These features allow facile recognition in sequenced genomes via directed genome-mining or using targeted PCR screens in environmental isolates (6, 16, 33, 34). The ready availability of PTM cluster sequences has empowered numerous biosynthetic inquiries via classical genetic, biochemical, and synthetic-biological (6, 8, 14–16, 20, 22, 24, 34–38) approaches. As a result, several gross chemical features of PTMs (their polycycle patterns, in particular) are now largely predictable by BGC ORF content and gene identity (17). In brief, PTM

cyclization cascades are driven by two classes of BGC-linked enzymes: flavin-dependent phytoene dehydrogenase homologs (PhyDHs) that form the A- and B-rings (e.g., IkaB, FtdC/FtdD, SGR813/SGR812, PtmB1/PtmB2, or OX1/OX2/OX3 (2, 8, 15, 32, 34)) and whenever present, nicotinamide-cofactor-dependent alcohol dehydrogenase (ADH) homologs (e.g., IkaC, FtdE and OX4) that form C-rings (39). Together, these two groups of enzymes are determinative for differentiating PTMs into polycycle type-classes. Illustrative PTM examples (Fig 1A) include combamide A with a 5/5 ring system (requiring two PhyDH homologs) (30), lysobacteramide B having a 5/5/6 ring system (requiring two PhyDHs and an ADH homolog) (15), and ikarugamycin-like chlokamycin which has a 5/6/5 ring system (requiring a single PhyDH plus an ADH homolog) (25). Finally, genes encoding cytochrome P450s (e.g., *ftdF*, *cftA*, *ikaD*, etc.) (2, 20, 26) and PTM C-25 oxidases (*ftdA* and homologs) (2, 14) are frequently associated with PTM BGCs where they direct oxidative decorations that diversify the class.

Despite the existence of >200 publications on PTM chemistry and biosynthesis, there remains sustained research interest in these molecules. This interest stems from several unusual features that have become hallmarks of the PTM family, including their atypical commonality and conservation across bacterial groups, the relative ease of their (bio)synthetic manipulation, their diverse bioactivities, plus their atypical enzymology. While core PTM cluster architectures are fairly conserved, an increasing number of PTM tailoring genes have been found outside the confines of their *ftdB*-centered BGCs (26, 40), which complicates product prediction. Finally, some PTM BGC-associated genes have biosynthetic functions that remain poorly predicted from sequence analysis alone (Fig 1B-C). Cytochrome P450s can perform not only regiodiverse oxidation chemistries,

but some (e.g., IkaD & CftA) also catalyze multiple oxidation events, leading to unexpected shunt paths (Fig 1C) (6, 20, 26). Some PTM phytoene dehydrogenases are multifunctional too; certain B-ring forming PhyDH enzymes (e.g., FtdC, OX1/OX2) catalyze oxy-insertions while others (e.g., CmbB, PtmB1) apparently do not (Fig 1B) (35). Finally, several other PTM molecular features lack known enzymology, including the chlorine decorations of chlokamycin (18, 25), the terminal olefin of frontalamides (2), the apparent loss of backbone carbons in cylindramide and geodin (41, 42), and the oxidative decoration of PTMs whose BGCs lack CYP450s (18). Finally, because PTM BGCs continue to be encountered in newly sequenced genomes, frequent reassessment of the class becomes necessary.

Here, we describe the systematic classification of PTM BGCs from 302 bacterial strains to document the diversity and distribution of PTM BGCs. Our phylogenetic and genome mining efforts revealed new PTM cluster architectures are discoverable from local soil actinomycetes and indicate most PTM BGCs are vertically transmitted. Our gene-by-gene deconstructive approach was able to detect “hidden” diversity among seemingly syntenic BGCs, and in certain clusters it revealed surprising mosaicism that hints towards infrequent but important horizontal gene transfer. Finally, we demonstrated that predicting biosynthetic outcomes via bioinformatics alone remains challenging. This followed from BGC gene-reduction experiments in *Streptomyces* sp. strain JV180 and *Streptomyces olivaceus* NRRL B-3009, plus the comparison of several other wild strains. This work provides a useful map of systematically classified PTM clusters, while our PTM-targeted metabolomics exposes limitations in using BGC data to predict product diversity.

RESULTS AND DISCUSSION

Identifying PTM BGCs in previously-sequenced and newly-isolated environmental bacteria. Our searches for PTM BGCs focused on detecting FtdB orthologs because the enzyme is essential for all PTM biosyntheses, its polyketide synthase functions are different from all known type-I iterative PKSs, and FtdBs are readily differentiable from non-FtdB PKS-NRPSs *in silico* (6, 16, 33, 34). The relatively large size of FtdBs (3069-3354 aa; Table S1) lend themselves to well-resolved phylogenetic analysis (Fig 2, Fig S1). Our targeted searches to locate PTM BGCs in public repositories (see Methods) yielded 284 BGCs (after manual dereplication to prevent duplicate analysis, Table S2).

We surmised some PTM BGC types would be rare or under-represented in public databases. To create a more diverse study set, we leveraged an *ftdB*-targeted degenerate-PCR screen (6) to identify PTM BGCs in actinobacterial isolates sourced from local soils. Based on amplicon sequence divergence or perceived scarcity, we selected 18 strains for complete cluster analysis via whole-genome sequencing (see nucleotide accessions). Along with complete PTM BGCs, this yielded *atpD*, *recA*, *rpoB*, and *trpB* sequences for host multi-locus (MLSA) phylogeny. Thus, public and environmental sequences yielded 302 complete PTM BGCs for comparison.

In terms of producer phylogeny, these efforts immediately expanded the catalog of putative PTM producers into the actinobacterial genera *Actinophytocola*, *Amycolatopsis*, *Thermomonospora*, and *Streptosporangium*. Several new proteobacterial producers were also identified including *Chitinimonas*, plus several metagenome assembled genomes (MAG) including *Cellvibrio* from citrus rhizosphere (43),

Thiohalocapsa from a 'pink berry' sulfur cycling consortium (44–46), and a putative Myxococcales bacterium found in wastewater sludge (47) (Fig S2).

A deconstructive approach to PTM BGC cataloging. Since the first PTM BGCs were sequenced >10 years ago, efforts to categorize their BGCs have been largely descriptive, guided mostly by synteny and the presence/absence of key biosynthetic genes. In pursuing a systematics-driven analysis of PTM BGC diversity, we assumed synteny-based binning of BGCs would eventually become limiting. This is because certain PTM-linked enzymes (especially cytochrome P450s and phytoene dehydrogenases) catalyze diverse reactions and are thus key drivers of PTM structural diversification (especially in combination). However, these enzymes are difficult to assign accurate functions to, especially in the absence of individual enzyme phylogenies. Accordingly, organizing PTM BGCs into discrete cluster types required particular attention to associated CYP450s (6, 20, 22, 26) and PhyDHs (8, 15, 18, 24, 34, 35, 48). We also avoided the use of concatenated trees (sometimes used to characterize other antibiotic BGCs (49, 50)) due to the high risk of conflating similarly organized but functionally different PTM loci. Thus, maximum-likelihood (ML) phylogenies of individual PTM-encoded enzymes (Fig 3, Fig S1, S3-5) were used for accurate, clade-based systematic binning of PTM enzymes and the resulting enzyme histories were used in-composite to identify meaningful enzyme sequence differences and evidence for BGC mosaicism.

PTM tailoring enzyme phylogenies. PTM-associated phytoene dehydrogenase-like (PhyDH) enzymes are named for homologs in carotenoid biosynthesis (CrtI (51)), but despite their specialized roles in PTM polycycle formation (34) (Fig 1B), the literature has lacked a standard, function-based nomenclature for these enzymes. Instead of continuing

historical naming conventions (e.g., OX1, OX2 and OX3; PhyDH1, PhyDH2, etc.), we adopted a phylogeny and function-based nomenclature rooted in the distinct catalyses of PTM PhyDH-like Ring-forming enzymes (Ppr's). Accordingly, the two PhyDHs required for the stepwise formation of a 5/5 ring PTM from FtdB products were termed PprA⁵s (initial cyclopentenyl, A-ring formation) and PprBs (for B-ring formation). In contrast, forming 5/6 polycycles (34) requires a single specialized PhyDH we termed PprA^{5/6}. Our Ppr ML phylogeny (Fig 3) multifurcates, with PprA⁵s forming discrete clades congruent with FtdB phylogeny. However, a distinct branch emerging from the Ppr polytomy reveals that all known PprA^{5/6} proteins (plus several synteny-suggested ones) are monophyletic. Interestingly, this suggests PprA^{5/6}s arose from ancestral PprA⁵s, suggesting 5/6/5 PTMs may have evolved after their 5/5/6 counterparts.

In contrast, PprBs branch deeply to form several subclades. Finding well-delineated PprB subclades initially suggested the possibility of ascribing precise catalytic function (oxygenation status, in particular) via conservation with literature- characterized members. However, we found non-hydroxylating and hydroxylating PprBs to be polyphyletic (Fig 3), preventing us from building on this idea and contrasting with earlier work in this vein (35). Our PhyDH analyses revealed other insights that may eventually inform catalysis. Most PprAs were distinguished from PprBs by a ~19 aa indel and Crtl homology suggests this region lies within a substrate-binding domain (51). This suggests a basis for differential substrate recognition and serves as useful landmark for parsing PprAs and PprBs from one another. However, we found intriguing exceptions to this pattern in PhyDHs extracted from PTM group 12c (Alteromonadales). Here, the associated PprAs and PprBs both contain a distinct sequence insertion from all other Pprs

(likely contributing to their PhyDH-clade resolution) (Fig 3). Interestingly, the 12c BGC *ppr* genes are arranged in reverse to all other known PTM BGCs (Fig 2). Recent analyses of Pel3, a 12c PprB, showed catalysis of a rare C6 hydroxylating, C7-C14 crosslinking B-ring variant during heterologous pseudoamide biosynthesis (48). These observations suggest an atypical evolutionary history for some Pprs, and the need for a more rigorous way of differentiating PTM PhyDHs beyond the indel region. Leveraging our PprA and PprB deep sequence comparisons, we revealed differentially enriched sequence motifs that should further assist classification efforts (Fig 3, Weblogo insets).

PTM-associated CYP450s have been intensely studied for understanding PTM oxidative decorations and exploiting the enzymes for molecular diversification. Our CYP450 phylogeny was created from an alignment of 232 CYP450s extracted from within the operon-like arrangements of core PTM BGCs, and those encoded nearby (within 10 kb) these clusters. The decision to include the latter group arises from increasing numbers of PTM-active enzymes (i.e., CYP450 IkaD (26) and HSAF *N*-methylase (40)) that locate outside of core BGCs, but are sometimes found nearby. Linking these unclustered CYP450s to PTM catalysis will require future experimentation so clade-numbers were not assigned to these particular CYP nodes (Fig S3). The resulting CYP450 ML phylogeny has two main branches, one comprised of CYP107-family enzymes (FtdFs) linked to Group 1 PTM BGCs (**I**, SGR and frontalamide-type PTMs) (2, 22, 23) plus multiple uncharacterized CYP450s upstream of Group 4 BGCs in certain *Kitasatospora*. The other main branch featured several subclades of CYP107 P450s with known biosynthetic roles, including CftA (**II**, clifednamides) (6, 20), IkaD (**VII.a**, ikarugamycin epoxide) (26), Mg1D (**III**, combamides B and G) (22), and CbmD (**VI**, combamides) (30). Other prominent

subclades featured characterized CYP105 family members (**IX, X**; SahD and SahE of sahamide biosynthesis) (10), yet- uncharacterized CYP450s likely acting in PTM biosynthesis due to gene clustering (**IV, V, VII.b, VIII**), plus BGC proximity-flagged ones (e.g., BE133 CYP(2)). Finally, some *Kitasatospora* CYP450s [*Kspr* CYP(2)] upstream of Group 4 BGCs were considerably diverged from other members of this branch.

FtdAs (PTM C-25 hydroxylases, Fig S4) and FtdEs (alcohol dehydrogenase-like C-ring forming enzymes, Fig S5) are critical for PTM structure determination and bioactivity (2, 39). In contrast to PhyDHs and cytochrome P450s, the catalyses performed by FtdAs and FtdEs are understood to be fairly conserved (14, 18). This suggests against them having pathway-specialized roles. The phylogenies of both enzymes appeared largely concordant with FtdB (Fig S1), suggesting sequence diversity seen in these proteins likely reflects drift and host speciation rather than catalytic diversity. This idea was supported by MLSA phylogeny of concatenated host housekeeping genes (Fig S2), where the most divergent FtdBs (an enzyme of highly conserved function) correlated with the most distant genera. Therefore, when interpreting PTM BGCs, we surmised *ftdA* and *ftdE* presence or absence should be emphasized over lower possibility of catalytic divergence.

A combination of FtdB phylogeny and BGC synteny is useful for estimating PTM diversity. Microbial genome-mining requires the accurate recognition of BGCs and identification of novel clusters. To do this for PTMs, a combination of evidence including traditional synteny, individual enzyme phylogenies and published PTM genetic, biochemical, and structural data were used where possible. BGC types identified using these data were anchored to an FtdB cladogram, resulting in a map that reflects both

sequence relatedness and differential enzyme content (Fig 2). Alpha-numeric identifiers were used to describe groups of BGCs according to their FtdB clades, and BGC variations within those clades. The first digit indicates FtdB branch (with the exceptions of Groups 3, 9 and 12, where groups of divergent-but-similar BGCs were binned). The letter code indicates plasticity (52), observed variability in gene content or arrangement within core BGC architectures. Decimals indicate the presence of potentially important tailoring genes encoded nearby (generally CYP450s).

PTM are divisible into two major families based on their polycycle patterns and the specialized enzymology associated with each. One is the 5/6/5 PTMs, and the other is comprised of the 5, 5/5, and 5/5/6 PTMs. Assigning BGCs to the first one was straightforward using the diagnostic presence of a PprA^{5/6} (but no PprB), plus an ADH. BGCs having these hallmarks were found in PTM groups 5, 6, and 11 (Fig 2) and distributed among diverse actinobacterial hosts. Group 5 clusters (5 a-c) were found in a distinct clade of *Streptomyces* spp (Fig S2) and include known loci for ikarugamycin and its oxidized analogs (capsimycins via IkaD, a type **VII.a** CYP450) (26). Interestingly, the *ikaDs* of Group 5 strains all locate downstream of their core ikarugamycin-type BGCs, but show highly variable distances between them (also observed in (37)). Group 6 BGCs were mostly found in *Kitasatospora* and *Streptomyces* thought to be plant-associated (Fig S2) (53–59). Their BGCs include several clifednamide loci (type **II** CYP450s, CftA) that include (6b) or lack (6a) *ftdAs* (6, 20). Interestingly, the *S. daliensis* group 6c BGC lacks a clustered *cftA*, thus likely producing ikarugamycin instead of clifednamides. Finally, the Group 11 BGCs from *Salinispora arenicola* and *Micromonospora* spp. (Fig S2) lack

271 CYP450s. Some members produce ikarugamycin (8) or butremycin (9), presumably via
272 cluster types 11b and 11a respectively.

273 In contrast with 5/6/5 PTMs, predicting the functions of the remaining PTM clusters
274 is more complex. 5/5/6 PTM carbocycles are produced stepwise, with ring A5 first (PprA⁵),
275 then B5 (PprB), then C6 (ADH). Thus, based on several biosynthetic studies (reviewed
276 (17)), BGCs lacking any one of these enzymes must direct lower (or no) polycycle
277 complexity. Thus, the simplest BGC architectures we encountered should produce
278 otherwise-acyclic tetramate polyenes similar to lysobacterene A (i.e., Group 3a (CYP450
279 adjacent), b (*ftdA*-adjacent), c (*ftdB*-alone)), molecules that are technically not polycyclic
280 macrolactams. Such compounds have never been found as native pathway products. As
281 with most BGCs analyzed here, it remains to be determined if these loci are active or are
282 past relics of genome rearrangement.

283 PTMs having a single five-membered (A5) carbocycle also remain undiscovered
284 from nature, but have been produced in engineered systems (e.g., lysobacterene B (15),
285 pactamide E (18, 24), sahamides A-D (10)). We identified potential native producers of
286 such compounds among the non-actinobacterial group 12 BGCs (Fig S2, Table S2).
287 *Saccharophagus degradans* (12a) and *Thiohalocapsa* sp. PB-PSB1 (12f) could produce
288 PprA-dependent A5 ring PTMs, while *Pelosinus fermentans* (12d) seems to encode an
289 unprecedented B5 PTM via its lone PprB.

290 5/5-polycyclic PTMs are produced by natural and pathway-engineered bacteria
291 (10, 12, 13, 15, 30, 48, 60, 61). In native strains they can arise from dedicated pathways,
292 but they are also associated with incomplete 5/5/6 PTMs biosynthesis (31, 32, 62, 63).
293 When mining for dedicated pathways, multiple clusters lacking adjacent ADH homologs

294 and encoding conserved PprA⁵ and PprBs were detected. *Streptomyces* group 2 BGCs
295 have a clustered *ftdA* and clade III CYP450, while the lone group 8 BGC encodes one
296 PprA and two PprB homologs. Group 8 *pprB* multiplicity is unusual and was here-to-fore
297 undocumented in actinomycetes, but it is similar to that of Gram-negative *Lysobacter*
298 *enzymogenes* C3 (Group 10a). In *L. enzymogenes*, PprB multiplicity drives 5/5/6
299 polycycle stereo-diversification (15), which could be tested in Group 8. Interestingly,
300 *Lysobacter antibioticus* (Group 10b) also appears to encode a dedicated 5/5 PTM
301 pathway; it lacks one of the PprB homologs and the ADH of *L. enzymogenes*. Multiple
302 other Gram-negative strains encode similar BGCs, including groups 12b and 12c. Finally,
303 Group 9 PTM BGCs of the Nocardiaceae and Pseudonocardaceae (Fig S2, Table S2)
304 encode several inferred 5/5 PTM pathways (*ftdA*-associated 9b, 9b.2; *ftdA*-minus 9e, 9f).

305 5/5/6 PTMs were well-documented from a diversity of bacterial producers in the
306 pre-genomic era (62, 64, 65). Several Group 1 BGCs have since been linked with 5/5/6
307 PTM production. Group 1 BGCs are exceptionally well-represented in genome
308 databases, comprising >50% of our dataset (Fig S1). Perhaps unsurprisingly, the Group
309 1 BGC architecture was among the first identified during early genome-mining efforts, and
310 5/5/6 PTMs (and certain biosynthetic intermediates) have been isolated from
311 *Streptomyces* (Fig S2) scattered across multiple Group 1 FtdB subclades. Known Group
312 1 PTM producers include *Streptomyces* sp. JV180 (maltophilin-like congeners) (38), *S.*
313 *griseus* IFO13350 (maltophilin-like, and alteramide-like congeners) (24, 32, 38),
314 *Streptomyces* sp. SCSIO40010 (10-epi-maltophilin) (23), multiple strains of *S.*
315 *albidoflavus* (somamycins, dihydromaltophilin [HSAF], 10-epi-HSAF, xanthobaccin C,

316 and alteramide stereoisomers) (19, 31, 66, 67), and *Streptomyces* sp. SPB78
317 (frontalamides) (2).

318 However, 5/5/6 PTM production is not limited to Group 1. Some bacteria having
319 groups 9a, 9c, and 10a BGCs produce 5/5/6 PTMs identical to certain Group 1 products
320 (7, 24, 65, 68). Interestingly, Group 9a, 9c, and 10a BGCs are similar in terms of gene
321 content but are distributed among highly divergent genera (*Streptomyces*,
322 *Actinoalloteichus*, *Saccharothrix*, *Actinokineospora*, and *Lysobacter*). Most of their known
323 products have a hallmark hydroxyl or carbonyl group on C14. Previous works suggest
324 hydroxylation at this position is PprB-catalyzed in Groups 1 and 10 (i.e., FtdC, OX1, OX2),
325 so we surmise analogous catalyses occur in Groups 9a and c (whose PprBs clade with
326 *Lysobacter* OX1 and OX2). However, the mechanism underlying C14 oxidation to a
327 carbonyl remains unknown.

328 *Streptomyces* and *Kitasatospora* Group 4 BGCs appeared similar to Group 1 5/5/6
329 clusters but lacked clade I CYP450s. Further, several members (Group 4a, a.2, c)
330 encoded a short gene (*ftdG*) for a Rieske 2Fe-2S-cluster protein, but this was not
331 definitive for the group (4b.2, 4c.2, 4c.3). Group 4 PTM production has never been
332 examined in native strains, but reconstitution of 4a BGC (excluding *ftdG*) from
333 *Streptomyces* sp. SCSIO 02999 in *Streptomyces lividans* TK64 resulted in the discovery
334 of 5/5/6 pactamides (18) and revealed the existence of C12-non-hydroxylating PprBs.
335 This critically differentiates SCSIO 02999 core PTM biosynthesis from other 5/5/6 PTMs
336 (18). Because Group 4 BGCs lack obviously clustered CYP450s but some pactamides
337 feature oxidative decorations on their polycycles (18), the need for unclustered oxidases
338 is implicated. This highlights CYP450s in the genome neighborhoods of BGCs 4a.2, 4b,

4b.2, 4c.2, 4c.3 for possible involvement (Fig S6). Finally, the biosynthetic role of *ftdG* in Group 4 BGCs remains enigmatic. Initial studies disregarded it as belonging to adjacent xiamycin BGCs (18, 69), but in SCSIO 02999 we noted the start codon of *ftdG* likely overlaps with the stop codon of upstream *adh*, which is common in operons. Furthermore, Group 4a.2 *Kitasatospora* BGCs feature *ftdGs* not juxtaposed with any xiamycin locus, again suggesting *ftdGs* are PTM- associated.

Group 7 PTM BGCs show unusual BGC diversification and mosaicism. Group 7 stood out against all other PTM BGC groups in our analysis because it includes five different BGC architectures linked to four different carbocyclic ring patterns (5, 5/5, 5/6/5 and 5/5/6; Fig 4). Group 7 BGCs are carried by several taxonomically disparate strains of *Streptomyces* (Fig S2), and our use of individual protein phylogenies revealed evidence of unusual mosaicism in the group. For example, group 7 FtdAs (where present) and FtdBs clade similarly between their respective ML phylogenies, but the *PprA*^{5/6} genes of Groups 7b and 7c appear most closely related to those of group 11 from Micromonosporaceae (Fig 3), suggesting gene flow from outside of the *Streptomyces* genus contributed group 7 BGCs for 5/6/5 PTMs. In contrast, Groups 7a and 7d harbor *pprA*⁵ and *pprBs*, predicting 5/5 and 5/5/6 PTM products, while the sole Group 7e BGC from *S. paludis* contains only a *pprA*⁵, thus putatively encoding PTMs with an A5 ring PTM. Interestingly, group 7a (5/5/6 carbocycles) and 7b-c (5/6/5 carbocycles) FtdEs are most related to those of other 5/6/5 clusters (IkaC- branch, Fig S5), again highlighting atypical gene patchworks in these BGCs.

There was also remarkable plasticity in the CYP450 gene positions and their assigned clades in group 7. The CYP450 (**IV** clade) of group 7a CYP450 is encoded

directly downstream of *ftdB*, a gene arrangement usually associated with 5/6/5
clifednamide BGCs (Group 6a/b). In turn, this unusual cluster outwardly appeared
identical to the Group 9d BGC of *Amycolatopsis sulphurea* (Fig 2), but these clusters are
non-orthologous because the latter has a clade **VIII** CYP450 (Fig S3). Group 7a and 9b
cluster architectures are both new to the PTM literature, with the former being discovered
in a streptomycete isolated from our local soils. In contrast with 7a, the remaining group
members have CYP450s (clades **V** and **VI**) encoded downstream of their core
biosynthetic genes. Most of the BGCs in group 7 remain experimentally uncharacterized,
but the *Streptomyces* sp. S10 group 7d cluster was interrogated via synthetic biology to
produce combamides (5/5 PTMs oxidatively tailored via clade **VI** CYP450 CmbD, Fig 1A,
and Fig S3 (30)).

PTM biosynthesis appears to have a complex history. Vertical inheritance is
argued to be the main mechanism of natural product BGC evolution, because on
evolutionary timescales horizontal gene transfer (HGT) of large clusters is likely
infrequent (70). However, it is also recognized that BGCs and the genes that comprise
them can have individual evolutionary paths (even within BGC families (71)). Some PTM
BGC types showed clear patterns of vertical transmission. For example, Group 5 clusters
were found only within a particular monophyletic clade of *Streptomyces* spp (Fig S2).
However, for many other PTM groups, such clear host -BGC concordance was absent.
Furthermore, our Group 7 analysis provided compelling evidence for HGT, albeit for a
subset of genes (Fig 4) rather than a complete BGC. Among the Group 7 BGCs is a
putative ikarugamycin cluster that, remarkably, may have arisen from gene exchange into
an ancestral 5/5/6 locus (based on nearest neighbors). It seems ikarugamycin BGCs may

be particularly subject to HGT, owing to their broad stochastic distribution among actinomycete hosts and their linkage with multiple FtdB clades (Fig 2). Finally, we noticed that at least ten PTM BGCs are adjacent to transposase genes, including members of Group 1, 4, 6, 9, and 12. It is intriguing as to what role these may have had in the spread of PTM BGCs. In sum, our analysis of FtdB and host phylogenies suggest an evolutionary history involving both vertical transmission and HGT.

New FtdB clade-specific probes are necessary capture PTM diversity. While less common since the advent of inexpensive genome sequencing, degenerate PCR remains a powerful approach for screening environmental microbes for genes and pathways of interest. We have isolated several PTM producing actinobacteria this way, using primers YQ158/YQ159 that target *ftdB*s ((6) and this study). However, those primers were designed when far fewer FtdB sequences were available and based on the much larger FtdB sequence library presented here, we realized new tools would be required to access the full diversity FtdB sequences. Accordingly, we used a phylogeny-guided degenerate oligonucleotide design strategy (see Methods, Table S3) and the resulting primers were validated against a panel of PTM⁺ bacteria having diverse FtdB sequences (Fig S7). This resulted in new primer sets that facilitated more robust cross-clade *ftdB* amplification than YQ158/159 (i.e., MTftdB1B, MTftdB4, MTftdB3B), while others gave robust amplification of certain *ftdB* sequences that were difficult to amplify with other primers sets (such as MTftdB6 for *Streptomyces* sp. SP18CS02 and MTftdB7 for *Pseudoalteromonas luteoviolacea*). Together, these new tools embody a robust toolset for the targeted discovery of PTM BGCs from environmental bacteria. Finally, sequence fragments of BGCs can be used to extrapolate potential BGC novelty, which

greatly empowers genome mining (i.e., KCDA genes in (72)). To test if the sequences of our new *ftdB* amplicons might predict PTM BGC novelty, a bank of degenerate-primer delimited amplicons was created in silico, and their in-frame translation products were used to build a ML phylogeny (Fig S8). This “amplicon tree” largely recapitulated the topology of the full-length FtdB tree (Fig S1, Fig S8), with some expected loss of resolution. Thus, FtdB amplicon phylogeny can be used to estimate PTM BGC novelty, within the constraints revealed by our FtdB-BGC concordance (Fig 2).

Using Group 1 PTM-targeted metabolomics to assess the fidelity of gene-to-molecules prediction. We previously showed actinobacterial strains harboring identical PTM clifednamide BGCs can display a range of congener titers, distributions, and types (6). Group 1 PTM BGCs had yet to be systematically surveyed in this way, and we thus asked how well PTM locus conservation translates into conserved product profiles, especially given how much FtdB diversity is represented within Group 1. To sample PTM Group 1 product diversity across species, we leveraged established PTM-targeted metabolomics to compare five strains that span the group’s FtdB diversity: *Streptomyces* sp. JV180, *Streptomyces* sp. SPB78, *Streptomyces* sp. JV186, *Streptomyces* sp. JV190, and *Streptomyces clavuligerus* ATCC 27064. Organic extracts from spent culture media of each were analyzed via high pressure liquid chromatography-tandem mass spectrometry (HPLC-MS/MS), with precursor ion scans for m/z 139.2 or 154.2 differentiating 25-deOH and 25-OH PTM congeners (38). The m/z transitions for each putative PTM are listed in Table S4, and product ion scans were performed for the most abundant base masses. The resulting MS² spectra were subsequently used to differentiate mass-similar features via fragment signatures, and to compare these against

known PTMs to assign membership to the PTM family. All putative PTMs were orthogonally identified by comparing product profiles against published Group 1 null mutants of *Streptomyces* spp strains SPB78 and JV180.

We found all tested group 1 strains produced highly similar PTM profiles (Fig 5, Fig S11A). Specifically, the tested strains all produced a PTM (**1**) that is possibly identical to biosynthetic intermediate FI-2 of SPB78 based on retention time, mass transition, and fragmentation pattern (Fig S9), plus another PTM (**13**) had high similarity to a maltophilin authentic standard (Fig S10, Fig S11A). In addition, several other commonly distributed putative PTM signals (**2**, **5**, **6**, **8**, **10**, **12-16**) were revealed by comparison against PTM null mutants. Furthermore, two minor peaks were shared by only a subset of strains: **3** in JV180, JV186, and JV190, and **11**, found in SPB78 and ATCC 27064. Finally, SPB78 generated multiple minor PTM signals not seen in other strains (**7**, **9**).

FtdF- type CYP450s are active when their cognate genes from *S. griseus*, *S. albidoflavus*, and *Streptomyces* sp. SPB78 are heterologously expressed (22), so we expected PTM structural consequences for *S. clavuligerus*, which lacks a typical Group 1 clustered *ftdF* (clade I CYP450). To probe why the latter strain still produces typical Group 1 PTM mass features, we first explored *ftdF*-dependent biosynthesis in *Streptomyces* sp. JV180, a genetically amenable host. This was done by creating two mutants lacking cluster-linked oxidative genes: *Streptomyces* sp. JV180 Δ *ftdA* (JV1520) and double mutant Δ *ftdA* Δ *ftdF* (JV1522) (Fig 5). Mutants in *ftdA* give simplified congener diversity (2), which increases intermediate detection sensitivity by reducing congener complexity. We thus revealed compounds **1**, **4**, **5**, **8** are *ftdF*-dependent, and these were clearly present in *S. clavuligerus* growth extracts (Fig 5). This indicated the *S. clavuligerus*

genome must harbor a functional FtdF ortholog, so BLAST was used to search the *S. clavuligerus* genome for FtdF- similar proteins along with CYP450 phylogeny to further interrogate potential hits (Fig S3). One CYP450 (69% identity to FtdF_{JV180}; EFG04254.1/FtdF_{ATCC27064}) stood out as a potential ortholog, which was encoded on the 1.8MB linear plasmid pSCL4 (73).

EFG04254.1 was cloned to constitutive-expression vector pJMD3 and tested for its ability to rescue *ftdF*-dependent mass features in JV180 $\Delta ftdA\Delta ftdF$ (in parallel with a FtdF_{JV180} control). After analyzing the complemented strains (JV2997 and JV3010, respectively), we found both CYP450 genes were functional in restoring production of **1** (FI-2), **4**, **5**, and **8** (Fig 5). The major restored mass feature, **1**, was further scrutinized to confirm its equivalency to products of parent strains via LC/MS and MS/MS fragmentation (Fig S9). JV3010 gave low production of **2** compared to the $\Delta ftdA$ and WT strains, but this may be a function of strong-constitutive expression (6). Linking EFG04254.1 to native *S. clavuligerus* PTM biosynthesis remains to be tested, but our heterologous rescue data show it catalyzes the appropriate reaction, making it a high-confidence candidate.

Group 1 and Group 4 comparative- minimization experiments expose weaknesses in synteny-based molecule prediction. The mutational analysis of JV180 Group 1 biosynthesis revealed product congener profiles can be effectively “collapsed” to a single prominent mass-product after deleting *ftdA* and *ftdF*. We questioned what would happen if we performed a parallel analysis with other gene clusters sharing the same core BGC architecture with Group 1. Would we recover the same biosynthetic intermediate? The Group 4 BGC of *Streptomyces olivaceus* NRRL B-3009 was selected to answer this because its core BGC outwardly matches JV180, with the exception that the Group 4

strain encodes a predicted Rieske 2Fe-2S protein instead of a FtdF-type CYP450. *S. olivaceus* NRRL B-3009 PTMs were previously unanalyzed, but our PTM-targeted metabolomics detected likely products via precursor ion scans, typical UV absorbance profiles, and certain PTM-type MS product ions (Fig 6, Fig S11B-E, Fig S12). As expected for strains having distinct BGC types, the putative PTM profiles of JV180 and B-3009 did not significantly overlap (Fig 6).

To create simplified PTM profiles in *S. olivaceus*, unmarked $\Delta ftdA$ and $\Delta ftdA\Delta ftdG$ mutants were created (JV1521 and JV1560) and analyzed. Like JV180, *ftdA* mutants were expected to cease 25-OH PTM production, which was evidenced in the complete loss of m/z 154.2 product ion signals (Fig S11C). Surprisingly, we observed no differences in the m/z 139.2 precursor profiles for either mutant when compared to parent controls (Fig 6), indicating *ftdG* may not contribute to PTM biosynthesis. This was further investigated via a $\Delta ftdG$ single mutant (JV1208), and like the other mutants, this strain also lacked obvious MS/MS profile differences compared to parent controls (Fig S11B). We thus concluded *ftdG* is inessential for PTM biosynthesis under the conditions tested; any function remains unknown.

The PTM intermediates produced by the “tailoring- minimized” JV180 $\Delta ftdA\Delta ftdF$ and B-3009 $\Delta ftdA\Delta ftdG$ strains were then compared via LC/MS/MS, and like their wild-type parents, these mutants again failed to produce overlapping profiles (Fig 6). This suggests the existence of yet-unrecognized contributors to PTM diversity, either hidden within the genes of the PTM BGCs themselves or in host factors outside of their cognate BGCs. Supporting the former idea, the NRRL B-3009 locus is similar to that for pactamides, molecules that lack C14 hydroxyl decorations (but feature other, yet-

unexplained oxidative modifications (18, 35)). Indeed, we detected a PTM signal (**31**) in NRRL B-3009 extracts which fragments consistently with a C14-deOH analog of **2** (Fig S11D-E, Fig S12A), perhaps identical to pactamide A (m/z 481.3). It is thus highly plausible that NRRL B-3009 utilizes a non-hydroxylating PprB, which would partially inform why the Group 1 and 4 “tailoring minimized” strains do not accumulate the same intermediate(s). However, this cannot completely explain the differences in product profiles; the B-3009 double mutant produces several other PTM-like mass features (**18-30**) in addition to **31**, but these too did not match any JV180 congeners (Fig 6, S11B, S12B-D), pointing to other mechanisms of pathway divergence. These data emphasize the need to disambiguate the complex origins of these molecules, especially in terms of extra-BGC enzymology (an increasingly encountered phenomenon (74–76)).

Lastly, the LC-MS/MS-based methods used here cannot resolve stereochemistry, a parameter with profound consequences on bioactivity. In light of recent PTM configuration revisions, Jin *et al* have argued for systematic re-investigation of PTM stereochemistry (24). Doing this would undoubtedly shed useful insights into structure, and eventually, refined bioinformatics models. Despite the limitations of MS, metabolomic studies on microbial chemistry are burgeoning (77–80). Because PTM BGCs are fairly common, their products are frequently encountered in these studies, but sometimes incorrectly. For example, specific PTMs have been extrapolated from MS signals whose structures are incongruent with the BGC content (i.e., butremycin from a Group 1 BGC) (81–83). Others indicate both 5/5/6 and 5/6/5 PTMs being produced by individual strains, which is unlikely without supporting genomic evidence (84–87). BGC information is thus urged as a priority dimension of these types of analyses.

METHODS

Strains, primers, plasmids, enzymes, chemicals, and general methods.

Strains used in metabolomics analysis are listed in Table S5; primers and plasmids used in strain construction are listed in Tables S6 and S7. All primers were purchased from Integrated DNA Technologies. All restriction enzymes were purchased from New England Biolabs. Taq polymerase and T4 Ligase were purchased from New England Biolabs or Thermo Fisher Scientific. Novagen KOD Hot Start polymerase was purchased from Millipore Sigma and used for all cloning. FailSafe PCR 2X PreMix D or G (LCG, Biosearch Technologies) were used as buffer for all PCR reactions.

Environmental actinomycete isolation and cultivation.

All bacterial isolates used in this study were enriched from a diversity of soils obtained from the Tyson Research Center of Washington University in St Louis, located in Eureka, MO. Isolates *Kitasatospora* sp. strains BE20 and BE303, and *Streptomyces* sp. strains BE133, BE147, BE230, BE282, and BE308 originated from New Rankin Cave sediments. *Kitasatospora* sp. strain SP17BM10 and *Streptomyces* sp. strain SP18ES09 were isolated from grassy rhizosphere soils of New Pond field. *Streptomyces* sp. SP17KL33 (6) and SP18CS02 were isolated from dry glade environments. Enrichment media used to obtain these strains were Humic-acid Vitamin (HV), Arginine Glycerol-Salts(AGS), and Low Tryptone-Yeast (LTY), and enrichments were performed as previously described using air-dried and finely sifted soils(33). All media contained antifungals cycloheximide and nystatin, (50mg/L or 100 mg/L ea) plus either nalidixic acid (25 mg/L) or polymyxin B (10 mg/L) to discourage non-actinomycete microorganisms. Specific enrichment media for each isolate were as follows: HV for strains BE133, BE147,

BE230, BE282, BE303, and BE308; AGS for strains BE20 and SP18CS02; and LTY for SP18ES09. Following isolation, the actinobacterial isolates were grown with vigorous shaking in 50mL conical tubes equipped with 5mL Trypticase Soy Broth and stored at -80°C for further use.

Draft genome sequencing of PTM BGC-harboring environmental isolates.

PTM⁺ environmental isolates above were identified using an established *ftdB*-targeted degenerate PCR screen (33). Screening amplicons were confirmed via Sanger sequencing at Genewiz using amplification primers YQ158/YQ159 (6). Parent strains of interest were draft sequenced using Illumina Genome sequencing at the Washington University in St. Louis McDonnell Genome Institute, with the TruSeq PCR-free library prep kit (Illumina) and the Sci clone next-generation sequencing (NGS) instrument (Perkin Elmer) essentially as previously reported (88).

PTM BGC mining. AntiSMASH software versions 3, 4, and 5 were used to identify PTM BGCs among fragmentary assembled genomes from Blodgett Lab environmental isolates (89). PTM BGCs were also systematically mined from online sequence databanks including GenBank (National Center for Biotechnological Information, NCBI) and the Joint Genome Institute Integrated Microbial Genomes and Microbiomes (JGI IMG) using BLASTP and TBLASTN tools. For this, the FtdB sequence from *Streptomyces* sp. JV178 (55) was used as a query against all available bacterial finished, permanent, and draft genome sequences. Hits above ~40% identity were selected as being potentially associated with PTM clusters, and PTM BGC identities were established through the identification of nearby encoded homologs of other known PTM BGC-associated genes. In addition, all candidate FtdB sequences

were tested for expected ornithine-specific adenylation domains using the PKS/NRPS Analysis website from The Institute for Genome Sciences at U of Maryland and NRPSsp Non-Ribosomal Peptide Synthase substrate predictor (90, 91). BGC linear comparisons were visualized using Clinker (92) where necessary.

Phylogenetic analyses of PTM BGC enzymes. For this study, the FtdB and cytochrome P450 trees were created with alignments of all available complete CDS translation products available from GenBank, plus those sequenced from environmental isolates. Specifically, these analyses included a total of 302 FtdBs and 232 CYP450s manual curated to remove exact duplicates. To ensure the quality of the FtdB tree, some FtdB having incomplete or faulty assembly were removed (including all <90% of typical FtdB length as determined by the average the lengths of 6 published FtdB's, plus 6 FtdB's sampling diverse genera found in this study (Table S1). In contrast, the FtdA, PhyDH, and ADH trees were created with enzyme sequences extracted from representative strains drawn from each BGC group (see Fig 2 for BGC group definitions). This was done for figure clarity and to high sequence redundancy. Sequences were aligned using Qiagen CLC Main Workbench v20. Maximum Likelihood trees were created using IQ-TREE2 with the model finder plus setting to determine the best-fitting model and 1000 ultra-fast bootstrapping replicates to assess branch supports. Ultrafast bootstrap threshold was set at 95%, and branches over 99% were bolded for each tree. Trees were visualized in CLC Main Workbench. The best-fit models selected for each tree are listed: FtdA: mtlInv+F+I+G4, FtdB: JTT+F+R10, FtdB amplicon: JTT+F+R10, Ppr: LG+F+R7, ADH: LG+F+R5, Cyp450: LG+F+R5.

Specific outgroups to root each phylogeny were: for PhyDH (IkaB/FtdC), carotene dehydratase family enzyme from *Rhodobacter capsulatus* (Crtl) (CAA77540.1) from (2); for cytochrome p450s, *Bacillus megaterium* BM3 (WP_095379236.1) from (6); for FtdB an iterative Type 1 PKS enzyme from *Saccharopolyspora erythraea* (WP_009944722.1) was used. To identify suitable outgroups to root ADH and FtdA trees, examples of each PTM associated protein were analyzed using BlastP (NCBI) to identify associated protein superfamilies. This revealed candidate non-PTM associated members for each superfamily. For the PTM ADH tree, an oxidoreductase from *Bacillus subtilis* subsp. *subtilis* str. 168 (CAB12574.1) was selected, and for FtdA, the root is a sterol desaturase from *Agrobacterium fabrum* str. C58 (Q7D1P6).

Taxonomic classification PTM BGC-harboring strains. PTM BGC-harboring genomes were downloaded from NCBI, as well as select partial housekeeping sequences (*atpD*, *recA*, *rpoB*, and *trpB*) of type strains used in previous taxonomic studies (93). Automlsa2 was used with model finder plus and 1000 ultrafast bootstrap replicates to mine and align sequences, and then build a phylogenetic tree (Best-fit model: GTR+F+R8:atpD+recA,GTR+F+R5:rpoB,TVM+F+R6:trpB) (94). CLC Main Workbench v20.0 was used to visualize alignments and trees. Published phylogenies using MLSA, whole genome analysis, and ANI scores were taken into consideration when analyzing taxonomic classification (93, 95–102).

PTM production and targeted metabolomics. Strains were revived from -80°C cryostocks onto ISP2 agar and incubated 2-3 d at 28°C. Agar plugs with adhering mycelia were used to inoculate single colonies to 3 ml TSB in 24-well plates. Cultures were shaken for 2 days at 28°C (250 rpm) with a sterile 4mm glass bead to disrupt

clumping. 200 μ l of each culture was spread to ISP2, ISP4, ISPS, SFM, or YMS agar plates and incubated at 28°C for 6 days. The agar was diced, immersed in ethyl acetate overnight, and solvent was evaporated under vacuum. The dry extracts were resuspended in 500 μ l LC-MS grade methanol and filtered before HPLC analysis. Chromatography was performed using a Phenomenex Luna C18 column (75 x 3 mm, 3 μ m pore size) installed on an Agilent 1260 Infinity HPLC. For each run, 10 μ l of sample was injected and the chromatography conditions were as follows: T = 0, 5% B; T = 1, 5% B; T = 4, 35% B; T = 22, 55% B; T = 26, 100% B; T = 30, 100%; A: water + 0.1% formic acid, B: acetonitrile + 0.1% formic acid; and 0.8 ml/min. An HPLC diode array detector (DAD) monitored PTM production at 320 nm, along with in-line Agilent 6420 Triple-Quad mass spectrometer for MS/MS analysis with ESI in positive mode (ESI+). For PTM-targeted metabolomics, precursor ion scan mode was used to detect molecules m/z 450 – 550 that fragmented (collision energy = 30 V, fragmentation energy = 70 V) into product ions m/z 139.2 and 154.2 as previously (38). Product ion scan mode was then used to detect precursor ions that were identified in the precursor scans, m/z : 493.2, 495.4, 497.4, 509.2, 511.2, 513.2, 525.2, 527.3, 529.0. MS data were analyzed offline using Agilent MassHunter Qualitative Analysis software, where PTMs were identified by correlating baseline-corrected UV profiles with PTM-like fragmentation patterns, established in (2, 6, 38).

Genetics: Marker-less gene deletion and rescue experiments Streptomycin counterselections were used to construct unmarked mutants to test gene function essentially as described (6, 38). Briefly, protein S12 (StrR) point mutants of *S. olivaceus* NRRL B-3009 were isolated by plating cells to ISP2 + Str¹⁰, and *rpsL* was amplified

from resistant colonies and sequenced using primers YQ4 and YQ5 (Table S7). The *rpsL* mutants used as parents for downstream mutation analyses were chosen based on their having had no visible difference in PTM production or growth characteristics from WT. All gene deletions were generated using double homologous recombination as previously described (6, 38). Deletion plasmids were generated by first cloning PCR-amplified homologous upstream and downstream regions into linearized pUC19 via NEB Hi-Fi Assembly and verified by Sanger sequencing. Homology cassettes were subcloned to pJVD52.1. Resulting plasmids were introduced into *Streptomyces* via intergeneric conjugation using JV36, where apramycin-resistant (Apr^R), Str^S exconjugants were selected. These were grown in TSB at 37°C without antibiotics, and double recombinants were identified via PCR after plating to ISP2 + streptomycin. Mutant rescue and heterologous expression were carried out using expression-shuttle vector pJMD3. Inserts were confirmed by Sanger sequencing using primers *Perme*^{*}-fw and PXS6 or via whole plasmid sequencing by Plasmidsaurus using Oxford Nanopore Technologies. Expression constructs were conjugated into recipients using JV36 by selecting for Apr^R colonies, essentially as previously described (2). Plasmid integration at Φ C31 *attB* was confirmed using multiplex PCR as done (6). See Table S6 for specific plasmid details.

FtdB Clade-based primer design and function tests. A total of 11 degenerate primer sets (see Table S3) were designed based *ftdB* nucleotide sequences drawn from distinct phylogenetic groups of the FtdB ML- tree (Fig S1). Degenerate primers were designed to specifically amplify across the junction of the PKS and NRPS modules of the enzymes using Primers4Clades (103). Primers were designed using the default *get*

660 *primers* run mode, and with corrected CODEHOP (codeh_corr) formulations. A set of
661 positive control strains (Table S3) were selected to validate the functionality of the new
662 degenerate primer pairs. *Streptomyces* sp. SP18CS02 has a newly-described cluster
663 architecture, and it was also used to test new primers because the strain saw
664 inconsistent amplification with YQ158/YQ159 (6). Universal 16S rDNA primers
665 8F/1492R (104) were used to control for template quality. All new degenerate *ftdB*
666 screening primers were used with the following “touchdown” protocol: 2 min at 95 °C,
667 followed by 7 cycles consisting of 20 s at 95 °C, 15 s at 70 °C, and 1 min at 70 °C,
668 decreasing annealing temperature by 2 °C after each cycle. This was followed by 23
669 cycles consisting of 20 s at 95 °C, 15 s at 54 °C, and 1 min at 68 °C, ending with 5 min
670 at 72 °C. All primer sets yielded expected amplicon sizes based on clade-specific
671 positive-control *ftdB* gene sequences (see Table S3).

672 **Nucleotide accession numbers.** The Whole Genome Shotgun Sequencing
673 data have deposited at DDBJ/ENA/GenBank under accessions: JAQYWT000000000
674 for *Streptomyces* sp. SP18BB07, JAQYWU000000000 for *Streptomyces* sp. BE133,
675 JAQYWV000000000 for *Streptomyces* sp. BE147, JAQYWW000000000 for
676 *Streptomyces* sp. BE20, JAQYWX000000000 for *Streptomyces* sp. BE230,
677 JAQYWY000000000 for *Streptomyces* sp. BE282, JAQYWZ000000000 for
678 *Streptomyces* sp. BE303, JAQYXA000000000 for *Streptomyces* sp. BE308,
679 JAQYXB000000000 for *Streptomyces* sp. SP17BM10, JAQYXC000000000 for
680 *Streptomyces* sp. SP18CS02, JAQYXD000000000 for *Streptomyces* sp. SP18ES09,
681 JAQYXE000000000 for *Streptomyces* sp. JV176, JAQYXF000000000 for *Streptomyces*
682 sp. JV181, JAQYXG000000000 for *Streptomyces* sp. JV184, JAQYXH000000000 for

683 *Streptomyces* sp. JV185, JAQYXI0000000000 for *Streptomyces* sp. JV186,
684 JAQYXJ0000000000 for *Streptomyces* sp. JV190, and JAQYXK0000000000 for
685 *Streptomyces* sp. SP17KL33.

686 SUPPLEMENTAL MATERIALS

687 Fig S1 to S9 and Table S1 to S7.

688 ACKNOWLEDGMENTS

689 We thank John D'Alessandro, Brandon Burger, Eric Song, Bilal Makhdoom, and Carrie
690 Stump for assisting strain isolations, Priya Mullick for cloning assistance, and Catrina
691 Fronick for draft genome assembly. For strains or DNA to validate degenerate PCR
692 toolsets, we thank: Paul Jensen (University of California- San Diego; *Salinispora*
693 *arenicola* CNS-205;), Liangcheng Du (University of Nebraska-Lincoln; *Lysobacter*
694 *enzymogenes* C3), and Avena Ross (Queen's University-Ontario; *Pseudoalteramonas*
695 *luteoviolacea* 2ta16). This work was supported by NSF-CAREER 1846005 to J.A.V.B.
696 Undergraduate researcher support is also gratefully acknowledged to: E.Z (Amgen
697 Scholars Award), A. D and M.T (Washington University Biology Summer Undergraduate
698 Research Fellowships). M.T also received a Lynda A. Ceremsak & F. George Davitt
699 Undergraduate Biochemistry Fellowship and a WUSTL-Howard Hughes Medical
700 Institute Summer Undergraduate Research Fellowship. J.A.V.B received genome-
701 sequencing support through a McDonnell Genome Institute 2018 Symposium Pilot
702 Project funded by the Office of the Dean of the School of Medicine, Washington
703 University Institute of Clinical and Translational Sciences (ICTS), and Illumina, Inc. (San
704 Diego, CA).

705 References:

- 706 1. Yu F, Zaleta-Rivera K, Zhu X, Huffman J, Millet J, Harris S, Yuen G, Li X-C, Du L.
707 2007. Structure and biosynthesis of heat-stable antifungal factor (HSAF), a broad-
708 spectrum antimycotic with a novel mode of action. *Antimicrob Agents Chemother*
709 51:64–72.
- 710 2. Blodgett JAV, Oh D-C, Cao S, Currie CR, Kolter R, Clardy J. 2010. Common
711 biosynthetic origins for polycyclic tetramate macrolactams from phylogenetically
712 diverse bacteria. *Proc Natl Acad Sci* 107:11692–11697.
- 713 3. Vior NM, Lacroix R, Chandra G, Dorai-Raj S, Trick M, Truman AW. 2018.
714 Discovery and biosynthesis of the antibiotic bicyclomycin in distantly related
715 bacterial classes. *Appl Environ Microbiol* 84:e02828-17.
- 716 4. Bown L, Li Y, Berru   F, Verhoeven JTP, Dufour SC, Bignell DRD. 2017.
717 Coronafacoyl phytotoxin biosynthesis and evolution in the common scab
718 pathogen *Streptomyces scabiei*. *Appl Environ Microbiol* 83:e01169-17.
- 719 5. Jiao Y-J, Liu Y, Wang H-X, Zhu D-Y, Shen Y-M, Li Y-Y. 2020. Expression of the
720 clifednamide biosynthetic pathway in *Streptomyces* generates 27,28-seco-
721 derivatives. *J Nat Prod* 83:2803–2808.
- 722 6. Qi Y, Ding E, Blodgett JAV. 2018. Native and engineered clifednamide
723 biosynthesis in multiple *Streptomyces* spp. *ACS Synth Biol* 7:357–362.
- 724 7. Xiangui M, Liping W, Dongyang W, Jie F, Weiming Z. 2017. Polycyclic tetramate
725 macrolactams from the marine-derived *Actinoalloteichus cyanogriseus* WH1-
726 2216-6. *Chin J Org Chem* 37:2352.

8. Greunke C, Glöckle A, Antosch J, Gulder TAM. 2017. Biocatalytic total synthesis of ikarugamycin. *Angew Chem Int Ed* 56:4351–4355.
9. Kyeremeh K, Acquah KS, Sazak A, Houssen W, Tabudravu J, Deng H, Jaspars M. 2014. Butremycin, the 3-hydroxyl derivative of ikarugamycin and a protonated aromatic tautomer of 5'-methylthioinosine from a Ghanaian *Micromonospora* sp. K310. *Mar Drugs* 12:999–1012.
10. Zou H, Xia X, Xu Q, Wang H, Shen Y, Li Y. 2022. Discovery of oxidized polycyclic tetramate macrolactams bearing one or two rings through combinatorial pathway reassembly. *Org Lett* 24:6515–6519.
11. Hoshino S, Wong CP, Ozeki M, Zhang H, Hayashi F, Awakawa T, Asamizu S, Onaka H, Abe I. 2018. Umezawamides, new bioactive polycyclic tetramate macrolactams isolated from a combined-culture of *Umezawaea* sp. and mycolic acid-containing bacterium. *J Antibiot (Tokyo)* 71:653–657.
12. Moree WJ, McConnell OJ, Nguyen DD, Sanchez LM, Yang Y-L, Zhao X, Liu W-T, Boudreau PD, Srinivasan J, Atencio L, Ballesteros J, Gavilán RG, Torres-Mendoza D, Guzmán HM, Gerwick WH, Gutiérrez M, Dorrestein PC. 2014. Microbiota of healthy corals are active against fungi in a light-dependent manner. *ACS Chem Biol* 9:2300–2308.
13. Ueoka R, Bhushan A, Probst SI, Bray WM, Lokey RS, Linington RG, Piel J. 2018. Genome-based identification of a plant-associated marine bacterium as a rich natural product source. *Angew Chem Int Ed* 57:14519–14523.

- 748 14. Greunke C, Antosch J, M. Gulder TA. 2015. Promiscuous hydroxylases for the
749 functionalization of polycyclic tetramate macrolactams – conversion of
750 ikarugamycin to butremycin. *Chem Commun* 51:5334–5336.
- 751 15. Li Y, Wang H, Liu Y, Jiao Y, Li S, Shen Y, Du L. 2018. Biosynthesis of the
752 polycyclic system in the antifungal HSAF and analogues from *Lysobacter*
753 *enzymogenes*. *Angew Chem Int Ed* 57:6221–6225.
- 754 16. Antosch J, Schaefer F, Gulder TAM. 2014. Heterologous reconstitution of
755 ikarugamycin biosynthesis in *E. coli*. *Angew Chem Int Ed* 53:3011–3014.
- 756 17. Zhang G, Zhang W, Saha S, Zhang C. 2016. Recent advances in discovery,
757 biosynthesis and genome mining of medicinally relevant polycyclic tetramate
758 macrolactams. *Curr Top Med Chem* 16:1727–1739.
- 759 18. Saha S, Zhang W, Zhang G, Zhu Y, Chen Y, Liu W, Yuan C, Zhang Q, Zhang H,
760 Zhang L, Zhang W, Zhang C. 2017. Activation and characterization of a cryptic
761 gene cluster reveals a cyclization cascade for polycyclic tetramate macrolactams.
762 *Chem Sci* 8:1607–1612.
- 763 19. Hou L, Liu Z, Yu D, Li H, Ju J, Li W. 2020. Targeted isolation of new polycyclic
764 tetramate macrolactams from the deepsea-derived *Streptomyces somaliensis*
765 SCSIO ZH66. *Bioorganic Chem* 101:103954.
- 766 20. Yang J, Qi Y, Blodgett JAV, Wencewicz TA. 2022. Multifunctional P450
767 monooxygenase CftA diversifies the clifednamide pool through tandem C–H bond
768 activations. *J Nat Prod* 85:47–55.

- 769 21. Ding W, Tu J, Zhang H, Wei X, Ju J, Li Q. 2021. Genome mining and metabolic
770 profiling uncover polycyclic tetramate macrolactams from *Streptomyces*
771 *koyangensis* SCSIO 5802. Mar Drugs 19:440.
- 772 22. Yan Y, Wang H, Song Y, Zhu D, Shen Y, Li Y. 2021. Combinatorial biosynthesis
773 of oxidized combamides using cytochrome P450 enzymes from different
774 polycyclic tetramate macrolactam pathways. ACS Synth Biol 10:2434–2439.
- 775 23. Liu W, Zhang W, Jin H, Zhang Q, Chen Y, Jiang X, Zhang G, Zhang L, Zhang W,
776 She Z, Zhang C. 2019. Genome mining of marine-derived *Streptomyces* sp.
777 SCSIO 40010 leads to cytotoxic new polycyclic tetramate macrolactams. Mar
778 Drugs 17:663.
- 779 24. Jin H, Zhang W, Zhang G, Zhang L, Liu W, Zhang C. 2020. Engineered
780 biosynthesis of 5/5/6 type polycyclic tetramate macrolactams in an ikarugamycin
781 (5/6/5 type)-producing chassis. Org Lett 22:1731–1735.
- 782 25. Fukuda T, Takahashi M, Kasai H, Nagai K, Tomoda H. 2017. Chlokamycin, a new
783 chloride from the marine-derived *Streptomyces* sp. MA2-12. Nat Prod Commun
784 12:1223–1226.
- 785 26. Yu H-L, Jiang S-H, Bu X-L, Wang J-H, Weng J-Y, Yang X-M, He K-Y, Zhang Z-G,
786 Ao P, Xu J, Xu M-J. 2017. Structural diversity of anti-pancreatic cancer
787 capsimycins identified in mangrove-derived *Streptomyces xiamenensis* 318 and
788 post-modification via a novel cytochrome P450 monooxygenase. Sci Rep
789 7:40689.

- 790 27. Malcomson B, Wilson H, Veglia E, Thillaiampalam G, Barsden R, Donegan S, El
791 Banna A, Elborn JS, Ennis M, Kelly C, Zhang S-D, Schock BC. 2016. Connectivity
792 mapping (ssCMap) to predict A20-inducing drugs and their antiinflammatory
793 action in cystic fibrosis. *Proc Natl Acad Sci* 113:E3725–E3734.
- 794 28. Jomon K, Kuroda Y, Ajisaka M, Sakai H. 1972. A new antibiotic, ikarugamycin. *J*
795 *Antibiot (Tokyo)* 25:271–280.
- 796 29. Jiang S-H, Dong F-Y, Da L-T, Yang X-M, Wang X-X, Weng J-Y, Feng L, Zhu L-L,
797 Zhang Y-L, Zhang Z-G, Sun Y-W, Li J, Xu M-J. 2020. Ikarugamycin inhibits
798 pancreatic cancer cell glycolysis by targeting hexokinase 2. *FASEB J* 34:3943–
799 3955.
- 800 30. Liu Y, Wang H, Song R, Chen J, Li T, Li Y, Du L, Shen Y. 2018. Targeted
801 discovery and combinatorial biosynthesis of polycyclic tetramate macrolactam
802 combamides A–E. *Org Lett* 20:3504–3508.
- 803 31. Olano C, García I, González A, Rodríguez M, Rozas D, Rubio J, Sánchez-Hidalgo
804 M, Braña AF, Méndez C, Salas JA. 2014. Activation and identification of five
805 clusters for secondary metabolites in *Streptomyces albus* J1074. *Microb*
806 *Biotechnol* 7:242–256.
- 807 32. Luo Y, Huang H, Liang J, Wang M, Lu L, Shao Z, Cobb RE, Zhao H. 2013.
808 Activation and characterization of a cryptic polycyclic tetramate macrolactam
809 biosynthetic gene cluster. 1. *Nat Commun* 4:2894.

- 810 33. Cao S, Blodgett JAV, Clardy J. 2010. Targeted discovery of polycyclic tetramate
811 macrolactams from an environmental *Streptomyces* strain. *Org Lett* 12:4652–
812 4654.
- 813 34. Zhang G, Zhang W, Zhang Q, Shi T, Ma L, Zhu Y, Li S, Zhang H, Zhao Y-L, Shi
814 R, Zhang C. 2014. Mechanistic insights into polycycle formation by reductive
815 cyclization in ikarugamycin biosynthesis. *Angew Chem Int Ed* 53:4840–4844.
- 816 35. Li X, Wang H, Li Y, Du L. 2019. Construction of a hybrid gene cluster to reveal
817 coupled ring formation–hydroxylation in the biosynthesis of HSAF and analogues
818 from *Lysobacter enzymogenes*. *MedChemComm* 10:907–912.
- 819 36. Xu L, Wu P, Wright SJ, Du L, Wei X. 2015. Bioactive polycyclic tetramate
820 macrolactams from *Lysobacter enzymogenes* and their absolute configurations by
821 theoretical ECD calculations. *J Nat Prod* 78:1841–1847.
- 822 37. Zhang W, Zhang G, Zhang L, Liu W, Jiang X, Jin H, Liu Z, Zhang H, Zhou A,
823 Zhang C. 2018. New polycyclic tetramate macrolactams from marine-derived
824 *Streptomyces* sp. SCSIO 40060. *Tetrahedron* 74:6839–6845.
- 825 38. Qi Y, Nepal KK, Blodgett JAV. 2021. A comparative metabologenomic approach
826 reveals mechanistic insights into *Streptomyces* antibiotic crypticity. *Proc Natl Acad*
827 *Sci* 118:e2103515118.
- 828 39. Li X, Wang H, Shen Y, Li Y, Du L. 2019. OX4 is an NADPH-dependent
829 dehydrogenase catalyzing an extended Michael addition reaction to form the six-
830 membered ring in the antifungal HSAF. *Biochemistry* 58:5245–5248.

- 831 40. Luo J, Li X, Wang H, Du L, Shen Y, Li Y. 2022. Identification and characterization
832 of the 28-*N*-methyltransferase involved in HSAF analogue biosynthesis.
833 *Biochemistry* 61:2879–2883.
- 834 41. Kanazawa S, Fusetani N, Matsunaga S. 1993. Cylindramide: Cytotoxic tetramic
835 acid lactam from the marine sponge *Halichondria cylindrata* Tanita & Hoshino.
836 *Tetrahedron Lett* 34:1065–1068.
- 837 42. Capon RJ, Skene C, Lacey E, Gill JH, Wadsworth D, Friedel T. 1999. Geodin A
838 magnesium salt: A novel nematocide from a southern Australian marine sponge,
839 *Geodia*. *J Nat Prod* 62:1256–1259.
- 840 43. Zhang Y, Xu J, Wang E, Wang N. 2020. Mechanisms underlying the rhizosphere-
841 to-rhizoplane enrichment of *Cellvibrio* unveiled by genome-centric metagenomics
842 and metatranscriptomics. *Microorganisms* 8:583.
- 843 44. Wilbanks EG, Jaekel U, Salman V, Humphrey PT, Eisen JA, Facciotti MT,
844 Buckley DH, Zinder SH, Druschel GK, Fike DA, Orphan VJ. 2014. Microscale
845 sulfur cycling in the phototrophic pink berry consortia of the Sippewissett Salt
846 Marsh. *Environ Microbiol* 16:3398–3415.
- 847 45. Wilbanks EG, Doré H, Ashby MH, Heiner C, Roberts RJ, Eisen JA. 2022.
848 Metagenomic methylation patterns resolve bacterial genomes of unusual size and
849 structural complexity. *ISME J* 16:1921–1931.
- 850 46. Kosmopoulos JC, Campbell DE, Whitaker RJ, Wilbanks EG. 2023. Horizontal
851 gene transfer and CRISPR targeting drive phage-bacterial host interactions and

coevolution in “pink berry” marine microbial aggregates. Appl Environ Microbiol 89:e00177-23.

47. Lin Y, Wang L, Xu K, Li K, Ren H. 2021. Revealing taxon-specific heavy metal-resistance mechanisms in denitrifying phosphorus removal sludge using genome-centric metaproteomics. Microbiome 9:67.

48. Li X, Liu Q, Zou H, Luo J, Jiao Y, Wang H, Du L, Shen Y, Li Y. 2023. Discovery and biosynthesis of pseudoamides reveal enzymatic cyclization of the polyene precursor to 5–5 bicyclic tetramate macrolactams. ACS Catal 13:4760–4767.

49. Joynt R, Seipke RF. 2018. A phylogenetic and evolutionary analysis of antimycin biosynthesis. Microbiology 164:28–39.

50. Chase AB, Sweeney D, Muskat MN, Guillén-Matus DG, Jensen PR. 2021. Vertical inheritance facilitates interspecies diversification in biosynthetic gene clusters and specialized metabolites. mBio 12:e02700-21.

51. Schaub P, Yu Q, Gemmecker S, Poussin-Courmontagne P, Mailliot J, McEwen AG, Ghisla S, Al-Babili S, Cavarelli J, Beyer P. 2012. On the structure and function of the phytoene desaturase CRTI from *Pantoea ananatis*, a membrane-peripheral and FAD-dependent oxidase/isomerase. PLoS ONE 7:e39550.

52. Claverys J, Prudhomme M, Mortier-Barrière I, Martin B. 2000. Adaptation to the environment: *Streptococcus pneumoniae*, a paradigm for recombination-mediated genetic plasticity? Mol Microbiol 35:251–259.

- 872 53. Gu L, Zhang K, Zhang N, Li X, Liu Z. 2020. Control of the rubber anthracnose
873 fungus *Colletotrichum gloeosporioides* using culture filtrate extract from
874 *Streptomyces deccanensis* QY-3. *Antonie Van Leeuwenhoek* 113:1573–1585.
- 875 54. Nguyen HP, Weisberg AJ, Chang JH, Clarke CR. 2022. *Streptomyces caniscabiei*
876 sp. nov., which causes potato common scab and is distributed across the world.
877 *Int J Syst Evol Microbiol* 72:005225.
- 878 55. Qi Y, D'Alessandro JM, Blodgett JAV. Draft genome sequence of *Streptomyces*
879 sp. strain JV178, a producer of clifednamide-type polycyclic tetramate
880 macrolactams. *Genome Announc* 6:e01401-17.
- 881 56. Guan D, Grau BL, Clark CA, Taylor CM, Loria R, Pettis GS. 2012. Evidence that
882 thaxtomin C is a pathogenicity determinant of *Streptomyces ipomoeae*, the
883 causative agent of *Streptomyces* soil rot disease of sweet potato. *Mol Plant-*
884 *Microbe Interactions* 25:393–401.
- 885 57. Planckaert S, Deflandre B, De Vries A-M, Ameye M, Martins JC, Audenaert K,
886 Rigali S, Devreese B. 2021. Identification of novel rothibin analogues in
887 *Streptomyces scabies*, including discovery of its biosynthetic gene cluster.
888 *Microbiol Spectr* 9:e00571-21.
- 889 58. Bown L, Hirota R, Goettge MN, Cui J, Krist DT, Zhu L, Giurgiu C, Van Der Donk
890 WA, Ju K-S, Metcalf WW. 2023. A novel pathway for biosynthesis of the herbicidal
891 phosphonate natural product phosphonothrixin is widespread in Actinobacteria. *J*
892 *Bacteriol* 205:e00485-22.

- 893 59. Armijos-Jaramillo V, Santander-Gordón D, Soria R, Pazmiño-Betancourth M,
894 Echeverría MC. 2017. A whole genome analysis reveals the presence of a plant
895 PR1 sequence in the potato pathogen *Streptomyces scabies* and other
896 *Streptomyces* species. Mol Phylogenet Evol 114:346–352.
- 897 60. Shigemori H, Bae MA, Yazawa K, Sasaki T, Kobayashi J. 1992. Alteramide A, a
898 new tetracyclic alkaloid from a bacterium *Alteromonas* sp. associated with the
899 marine sponge *Halichondria okadae*. J Org Chem 57:4317–4320.
- 900 61. Bae M-A, Yamada K, Ijuin Y, Tsuji T, Yazawa K, Uemura D. 1996.
901 Aburatubolactam A, a novel inhibitor of superoxide anion generation from a
902 marine microorganism. Heterocycl Commun 2:315–318.
- 903 62. Jakobi M, Winkelmann G, Kaiser D, Kempter C, Jung G, Berg G, Bahl H. 1996.
904 Maltophilin: A new antifungal compound produced by *Stenotrophomonas*
905 *maltophilia* R3089. J Antibiot (Tokyo) 49:1101–1104.
- 906 63. Tang B, Laborda P, Sun C, Xu G, Zhao Y, Liu F. 2019. Improving the production
907 of a novel antifungal alteramide B in *Lysobacter enzymogenes* OH11 by
908 strengthening metabolic flux and precursor supply. Bioresour Technol 273:196–
909 202.
- 910 64. Graupner PR, Thornburgh S, Mathieson JT, Chapin EL, Kemmitt GM, Brown JM,
911 Snipes CE. 1997. Dihydromaltophilin; a novel fungicidal tetramic acid containing
912 metabolite from *Streptomyces* sp. J Antibiot (Tokyo) 50:1014–1019.

- 913 65. Nakayama T, Homma Y, Hashidoko Y, Mizutani J, Tahara S. 1999. Possible role
914 of xanthobaccins produced by *Stenotrophomonas* sp. strain SB-K88 in
915 suppression of sugar beet damping-off disease. *Appl Environ Microbiol* 65:4334–
916 4339.
- 917 66. Quezada M, Licona-Cassani C, Cruz-Morales P, Salim AA, Marcellin E, Capon
918 RJ, Barona-Gómez F. 2017. Diverse cone-snail species harbor closely related
919 *Streptomyces* species with conserved chemical and genetic profiles, including
920 polycyclic tetramic acid macrolactams. *Front Microbiol* 8:2305.
- 921 67. Ahmed Y, Rebets Y, Tokovenko B, Brötz E, Luzhetskyy A. 2017. Identification of
922 butenolide regulatory system controlling secondary metabolism in *Streptomyces*
923 *albus* J1074. *Sci Rep* 7:9784.
- 924 68. Ding L, Zhang S-D, Haidar AK, Bajimaya M, Guo Y, Larsen TO, Gram L. 2021.
925 Polycyclic tetramate macrolactams – A group of natural bioactive metallophores.
926 *Front Chem* 9:772858.
- 927 69. Li H, Zhang Q, Li S, Zhu Y, Zhang G, Zhang H, Tian X, Zhang S, Ju J, Zhang C.
928 2012. Identification and characterization of xiamycin A and oxiamycin gene cluster
929 reveals an oxidative cyclization strategy tailoring indolosesquiterpene
930 biosynthesis. *J Am Chem Soc* 134:8996–9005.
- 931 70. Chevrette MG, Currie CR. 2019. Emerging evolutionary paradigms in antibiotic
932 discovery. *J Ind Microbiol Biotechnol* 46:257–271.

- 933 71. Chevrette MG, Gutiérrez-García K, Selem-Mojica N, Aguilar-Martínez C, Yañez-
934 Olvera A, Ramos-Aboites HE, Hoskisson PA, Barona-Gómez F. 2020.
935 Evolutionary dynamics of natural product biosynthesis in bacteria. *Nat Prod Rep*
936 37:566–599.
- 937 72. Shigdel UK, Lee S-J, Sowa ME, Bowman BR, Robison K, Zhou M, Pua KH, Stiles
938 DT, Blodgett JAV, Udvary DW, Rajczewski AT, Mann AS, Mostafavi S, Hardy T,
939 Arya S, Weng Z, Stewart M, Kenyon K, Morgenstern JP, Pan E, Gray DC, Pollock
940 RM, Fry AM, Klausner RD, Townson SA, Verdine GL. 2020. Genomic discovery of
941 an evolutionarily programmed modality for small-molecule targeting of an
942 intractable protein surface. *Proc Natl Acad Sci* 117:17195–17203.
- 943 73. Medema MH, Trefzer A, Kovalchuk A, van den Berg M, Müller U, Heijne W, Wu L,
944 Alam MT, Ronning CM, Nierman WC, Bovenberg RAL, Breitling R, Takano E.
945 2010. The sequence of a 1.8-MB bacterial linear plasmid reveals a rich
946 evolutionary reservoir of secondary metabolic pathways. *Genome Biol Evol*
947 2:212–224.
- 948 74. Losada AA, Cano-Prieto C, García-Salcedo R, Braña AF, Méndez C, Salas JA,
949 Olano C. 2017. Caboxamycin biosynthesis pathway and identification of novel
950 benzoxazoles produced by cross-talk in *Streptomyces* sp. NTK 937. *Microb*
951 *Biotechnol* 10:873–885.
- 952 75. Tsunematsu Y, Ishikawa N, Wakana D, Goda Y, Noguchi H, Moriya H, Hotta K,
953 Watanabe K. 2013. Distinct mechanisms for spiro-carbon formation reveal
954 biosynthetic pathway crosstalk. *Nat Chem Biol* 9:818–825.

- 955 76. Cano-Prieto C, Losada AA, Braña AF, Méndez C, Salas JA, Olano C. 2015.
956 Crosstalk of nataxazole pathway with chorismate-derived ionophore biosynthesis
957 pathways in *Streptomyces* sp. Tü 6176. ChemBioChem 16:1925–1932.
- 958 77. Avalon NE, Murray AE, Baker BJ. 2022. Integrated metabolomic–genomic
959 workflows accelerate microbial natural product discovery. Anal Chem 94:11959–
960 11966.
- 961 78. Kellogg J, Kang S. 2020. Metabolomics, an essential tool in exploring and
962 harnessing microbial chemical ecology. Phytobiomes J 4:195–210.
- 963 79. Mannocho-Russo H, Swift SOI, Nakayama KK, Wall CB, Gentry EC,
964 Panitchpakdi M, Caraballo-Rodriguez AM, Aron AT, Petras D, Dorrestein K,
965 Dorrestein TK, Williams TM, Nalley EM, Altman-Kurosaki NT, Martinelli M,
966 Kuwabara JY, Darcy JL, Bolzani VS, Wegley Kelly L, Mora C, Yew JY, Amend
967 AS, McFall-Ngai M, Hynson NA, Dorrestein PC, Nelson CE. 2023. Microbiomes
968 and metabolomes of dominant coral reef primary producers illustrate a potential
969 role for immunolipids in marine symbioses. Commun Biol 6:896.
- 970 80. Aharoni A, Goodacre R, Fernie AR. 2023. Plant and microbial sciences as key
971 drivers in the development of metabolomics research. Proc Natl Acad Sci
972 120:e2217383120.
- 973 81. Samples RM, Puckett SP, Balunas MJ. 2023. Metabolomics Peak Analysis
974 Computational Tool (MPACT): An advanced informatics tool for metabolomics and

975 data visualization of molecules from complex biological samples. Anal Chem
 976 95:8770–8779.

977 82. Pessotti R de C, Hansen BL, Reaso JN, Ceja-Navarro JA, El-Hifnawi L, Brodie
 978 EL, Traxler MF. 2021. Multiple lineages of *Streptomyces* produce antimicrobials
 979 within passalid beetle galleries across eastern North America. eLife 10:e65091.

980 83. Pinto-Almeida A, Bauermeister A, Luppino L, Grilo IR, Oliveira J, Sousa JR,
 981 Petras D, Rodrigues CF, Prieto-Davó A, Tasdemir D, Sobral RG, Gaudêncio SP.
 982 2021. The diversity, metabolomics profiling, and the pharmacological potential of
 983 actinomycetes isolated from the Estremadura Spur pockmarks (Portugal). Mar
 984 Drugs 20:21.

985 84. Paulus C, Rebets Y, Zapp J, Rückert C, Kalinowski J, Luzhetskyy A. 2018. New
 986 alpiniamides from *Streptomyces* sp. IB2014/011-12 assembled by an unusual
 987 hybrid non-ribosomal peptide synthetase trans-AT polyketide synthase enzyme.
 988 Front Microbiol 9:1959.

989 85. Armin R, Zühlke S, Mahnkopp-Dirks F, Winkelmann T, Kusari S. 2021. Evaluation
 990 of apple root-associated endophytic *Streptomyces pulveraceus* strain ES16 by an
 991 OSMAC-assisted metabolomics approach. Front Sustain Food Syst 5:643225.

992 86. Shaikh AA, Nothias L-F, Srivastava SK, Dorrestein PC, Tahlan K. 2021.
 993 Specialized metabolites from ribosome engineered strains of *Streptomyces*
 994 *clavuligerus*. 4. Metabolites 11:239.

- 995 87. Deutsch JM, Mandelare-Ruiz P, Yang Y, Foster G, Routhu A, Houk J, De La Flor
996 YT, Ushijima B, Meyer JL, Paul VJ, Garg N. 2022. Metabolomics approaches to
997 dereplicate natural products from coral-derived bioactive bacteria. *J Nat Prod*
998 85:462–478.
- 999 88. Qi Y, Nepal KK, Greif J, Martini C, Tomlinson C, Markovic C, Fronick C, Blodgett
1000 JAV. 2020. Draft genome sequences of two polycyclic tetramate macrolactam
1001 producers, *Streptomyces* sp. strains JV180 and SP18CM02. *Microbiol Resour*
1002 *Announc* 9:e01066-20.
- 1003 89. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T.
1004 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining
1005 pipeline. *Nucleic Acids Res* 47:W81–W87.
- 1006 90. Prieto C, García-Estrada C, Lorenzana D, Martín JF. 2012. NRPSsp: non-
1007 ribosomal peptide synthase substrate predictor. *Bioinformatics* 28:426–427.
- 1008 91. Bachmann BO, Ravel J. 2009. Chapter 8. Methods for *in silico* prediction of
1009 microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA
1010 sequence data, p. 181–217. *In* *Methods in Enzymology*. Elsevier.
- 1011 92. Gilchrist CLM, Chooi Y-H. 2021. clinker & clustermap.js: automatic generation of
1012 gene cluster comparison figures. *Bioinformatics* 37:2473–2475.
- 1013 93. Labeda DP, Dunlap CA, Rong X, Huang Y, Doroghazi JR, Ju K-S, Metcalf WW.
1014 2017. Phylogenetic relationships in the family Streptomycetaceae using multi-
1015 locus sequence analysis. *Antonie Van Leeuwenhoek* 110:563–583.

- 1016 94. Davis EW II, Okrent RA, Manning VA, Trippe KM. 2021. Unexpected distribution
1017 of the 4-formylaminooxyvinylglycine (FVG) biosynthetic pathway in *Pseudomonas*
1018 and beyond. PLOS ONE 16:1–20.
- 1019 95. Guo Y, Zheng W, Rong X, Huang Y. 2008. A multilocus phylogeny of the
1020 *Streptomyces griseus* 16S rRNA gene clade: Use of multilocus sequence analysis
1021 for streptomycete systematics. Int J Syst Evol Microbiol 58:149–159.
- 1022 96. Rong X, Huang Y. 2010. Taxonomic evaluation of the *Streptomyces griseus* clade
1023 using multilocus sequence analysis and DNA–DNA hybridization, with proposal to
1024 combine 29 species and three subspecies as 11 genomic species. Int J Syst Evol
1025 Microbiol 60:696–703.
- 1026 97. Labeda DP. 2011. Multilocus sequence analysis of phytopathogenic species of
1027 the genus *Streptomyces*. Int J Syst Evol Microbiol 61:2525–2531.
- 1028 98. 2012. Bergey's Manual of Systematic Bacteriology. Springer New York, New
1029 York, NY. Retrieved 3 March 2021.
- 1030 99. Labeda DP, Doroghazi JR, Ju K-S, Metcalf WW. 2014. Taxonomic evaluation of
1031 *Streptomyces albus* and related species using multilocus sequence analysis and
1032 proposals to emend the description of *Streptomyces albus* and describe
1033 *Streptomyces pathocidini* sp. nov. Int J Syst Evol Microbiol 64:894–900.
- 1034 100. Cheng K, Rong X, Pinto-Tomás AA, Fernández-Villalobos M, Murillo-Cruz C,
1035 Huang Y. 2015. Population genetic analysis of *Streptomyces albidoflavus* reveals

habitat barriers to homologous recombination in the diversification of streptomycetes. Appl Environ Microbiol 81:966–975.

101. Weisberg AJ, Kramer CG, Kotha RR, Luthria DL, Chang JH, Clarke CR. 2021. A novel species-level group of *Streptomyces* exhibits variation in phytopathogenicity despite conservation of virulence loci. Mol Plant-Microbe Interactions 34:39–48.

102. Rong X, Doroghazi JR, Cheng K, Zhang L, Buckley DH, Huang Y. 2013. Classification of *Streptomyces* phylogroup pratensis (Doroghazi and Buckley, 2010) based on genetic and phenotypic evidence, and proposal of *Streptomyces pratensis* sp. nov. Syst Appl Microbiol 36:401–407.

103. Contreras-Moreira B, Sachman-Ruiz B, Figueroa-Palacios I, Vinuesa P. 2009. primers4clades: a web server that uses phylogenetic trees to design lineage-specific PCR primers for metagenomic and diversity studies. Nucleic Acids Res 37:W95–W100.

104. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 1991. 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol 173:697–703.

Figure Legends

Fig 1. PTM structural diversity. (A) Example PTMs illustrating varying ring arrangements (brown numbers), tailored decorations (red & pink), and stereochemistry. (B) PTM PhyDHs catalyze an array of A and B ring arrangements. (C) PTMs sometimes involve multistep CYP450 oxidations, which are difficult to bioinformatically predict.

Fig 2. A FtdB- PTM BGC concordance. A cladogram representing 302 full-length FtdB sequences was collapsed for clarity and linked to a meta-analysis of associated PTM BGC types, representative strains, and CYP450 clades (see Figs S1 and Fig S3 for associated maximum-likelihood phylogenies). Note representative strains for each architecture are shown; this is not indicative of complete host taxonomy for each BGC type (See Fig S2 and Table S2 for details). Genus names with brackets indicate a potential taxonomic discrepancy with the current naming convention. Asterisks indicate strains examined via PTM-targeted metabolomics in this study.

Fig 3. PTM- associated PhyDH ML phylogeny and sequence analysis assist polycycle prediction. Branches are colored according to BGC groups, UF-Bootstrap threshold is 95%, and branches >99% are bolded. Stars denote experimentally characterized non-oxygenating PprBs. Protein domains are colored according to *Pantoea ananatis* Crtl homology (51), where green is FAD-binding, blue is substrate-binding, and orange is a non-conserved membrane helical region. Web logos illustrate amino-acid motifs differentially enriched in PprA or PprB-type PhyDHs. See Table S2 for abbreviations.

Fig 4. Group 7 BGCs are predicted to encode for an atypical diversity of PTM structures. FtdB protein phylogeny indicates the core biosynthetic enzymes of Group 7 BGCs are closely related, which starkly contrasts with the diverse tailoring gene content seen among individual BGC types. This suggests an unusual and exceptionally diverse array of PTM scaffolds are encoded by Group 7, compared to all other BGC groups identified in this work.

Fig 5. Group 1 strains produce consistent PTM congener profiles despite inconsistent *ftdF* arrangements. The 25-deOH PTM products (LC/MS/MS monitoring *m/z* 139.2 product ions) of multiple Group 1 BGC strains showed largely overlapping profiles compared to PTM⁻ mutant controls (in grey; JV168 and JV352). *ftdF*-dependent mass features were identified by comparing these against JV180 $\Delta ftdA$ and $\Delta ftdAftdF$ mutants (each with empty vector pJMD3; JV3003 and JV3006). These mutants and data were used to test for CYP450 orthology between an unclustered *S. clavuligerus* *ftdF* homolog (FtdF_{ATCC27064}, JV2997) and ectopic native *ftdF* (FtdF_{JV180}, JV3010). Asterisks indicate probable but very low abundance PTMs. The *m/z* transitions for numbered peaks are listed in Table S4.

Fig 6. Native and tailoring-minimized Group 4 (NRRL B-3009) and Group 1 (JV180) BGCs produce group-specific product profiles. LC/MS/MS monitoring of 25-deOH PTM profiles (*m/z* 139.2 product ions) produced by *Streptomyces* sp. JV180 (JV307), JV180 $\Delta ftdA$ (JV1520), JV180 $\Delta ftdA\Delta ftdF$ (JV1522), *S. olivaceus* NRRL B-3009 (JV1101), B-3009 $\Delta ftdA$ (JV1521), and B-3009 $\Delta ftdA\Delta ftdG$ (JV1560) revealed unique product profiles between the tested BGC groups (see text). Asterisks indicate PTMs were detected with certain MS/MS methods, but only in very low amounts. The *m/z* transitions for numbered peaks are listed in Table S4.