Structural and evolutionary analyses support reclassification of glycopeptide antibiotics as xyclopeptides

Short title: Reclassifying glycopeptide antibiotics

Athina Gavriilidou[1], Noel Kubach[1], Martina Adamek[1,2], Jens-Peter Rodler[3], Susanna Kremer[1], Daniel H. Huson[4], Rosa Alduina[5], Gerard Wright[6], Mohammad Seyemsayadost[7], Wolfgang Wohlleben[8,2], Stefano Donadio[9], Margherita Sosio[9], Min Xu[10,11], Max J. Cryle[12,13], Evi Stegmann*[2,3], Nadine Ziemert* [1,2]

1: Translational Genome Mining for Natural Products, Interfaculty Institute of Microbiology and Infection Medicine Tübingen (IMIT), Interfaculty Institute for Biomedical Informatics (IBMI), University of Tübingen, Tübingen, Germany / Cluster of Excellence 'Controlling Microbes to Fight Infections' (CMFI), University of Tübingen, Tübingen, Germany
2: German Centre for Infection Research (DZIF), Partnersite Tübingen, Tübingen, Germany
3: Microbial Bioactive Compounds, Interfaculty Institute of Microbiology and Infection Medicine Tübingen, University of Tübingen, Auf der Morgenstelle 28, 72076 Tübingen, Germany / Cluster of Excellence 'Controlling Microbes to Fight Infections' (CMFI), University of Tübingen, Tübingen, Germany
4: Algorithms in Bioinformatics, Computer Science Department, University of Tübingen, Tübingen, 72076, Germany, Interfaculty Institute for Biomedical Informatics (IBMI), University of Tübingen, Tübingen, Germany / Cluster of Excellence 'Controlling Microbes to Fight Infections' (CMFI), University of Tübingen, Tübingen / Germany
5: Department of Biological, Chemical and Pharmaceutical Sciences and Technologies (STEBICEF), University of Palermo, 90128 Palermo, Italy.
6: Department of Biochemistry & Biomedical Sciences, David Braley Center for Antibiotic Discovery, Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, Canada
7: Department of Chemistry, Princeton University, Princeton, New Jersey 08544, United States
8: Microbiology/Biotechnology, Interfaculty Institute of Microbiology and Infection Medicine Tübingen, University of Tübingen, Auf der Morgenstelle 28, 72076 Tübingen, Germany / Cluster of Excellence 'Controlling Microbes to Fight Infections' (CMFI), University of Tübingen, Tübingen, Germany
9:Naicons Srl, Milan, Italy
10: National Technology Innovation Center of Synthetic Biology, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China
11: Haihe Laboratory of Synthetic Biology, Tianjin 300308, China
12: Department of Biochemistry and Molecular Biology, The Monash Biomedicine Discovery Institute, Monash University, Clayton, VIC, 3800, Australia
13: ARC Centre of Excellence for Innovations in Peptide and Protein Science, Clayton, Victoria 3800, Australia
* co-corresponding author

Abstract

Glycopeptide antibiotics (GPAs) are key agents against multidrug-resistant Gram-positive pathogens, yet both the term “glycopeptide” and the current GPA type I-V classification framework have become increasingly strained as structurally and mechanistically divergent members continue to be discovered. In particular, compounds historically grouped as “type V GPAs” differ from classical GPAs in features such as glycosylation, peptide length, and reported mode of action, raising the question of whether they belong to the same natural product class. Here, a curated dataset of GPA-associated biosynthetic gene clusters (BGCs) is analysed by combining fingerprint similarity of the products with phylogenetic analysis of the BGCs. Fingerprint-based structural similarity networks and BGC similarity comparisons reveal a pronounced separation between classical lipid II-binding GPAs (types I-IV) and type V GPAs. Multi-locus phylogenetic analyses of conserved biosynthetic components further support two deeply divergent evolutionary subclasses, consistent with subclass-specific biosynthetic signatures. Together, these results motivate a revised, unambiguous framework in which the broader class is termed xyclopeptides, comprising the subclasses dalabactins (legacy GPA types I–IV) and murobactins (legacy type V).

Introduction

Glycopeptide antibiotics (GPA) have long been central to the treatment of infections caused by multidrug-resistant Gram-positive pathogens, exemplified by the discovery of vancomycin in 1953. Since then, dozens of natural products and numerous semi-synthetic derivatives have been described, several of which remain in clinical use. Accordingly, the biosynthesis and mechanisms of action of these compounds have been studied extensively over decades [1–9].

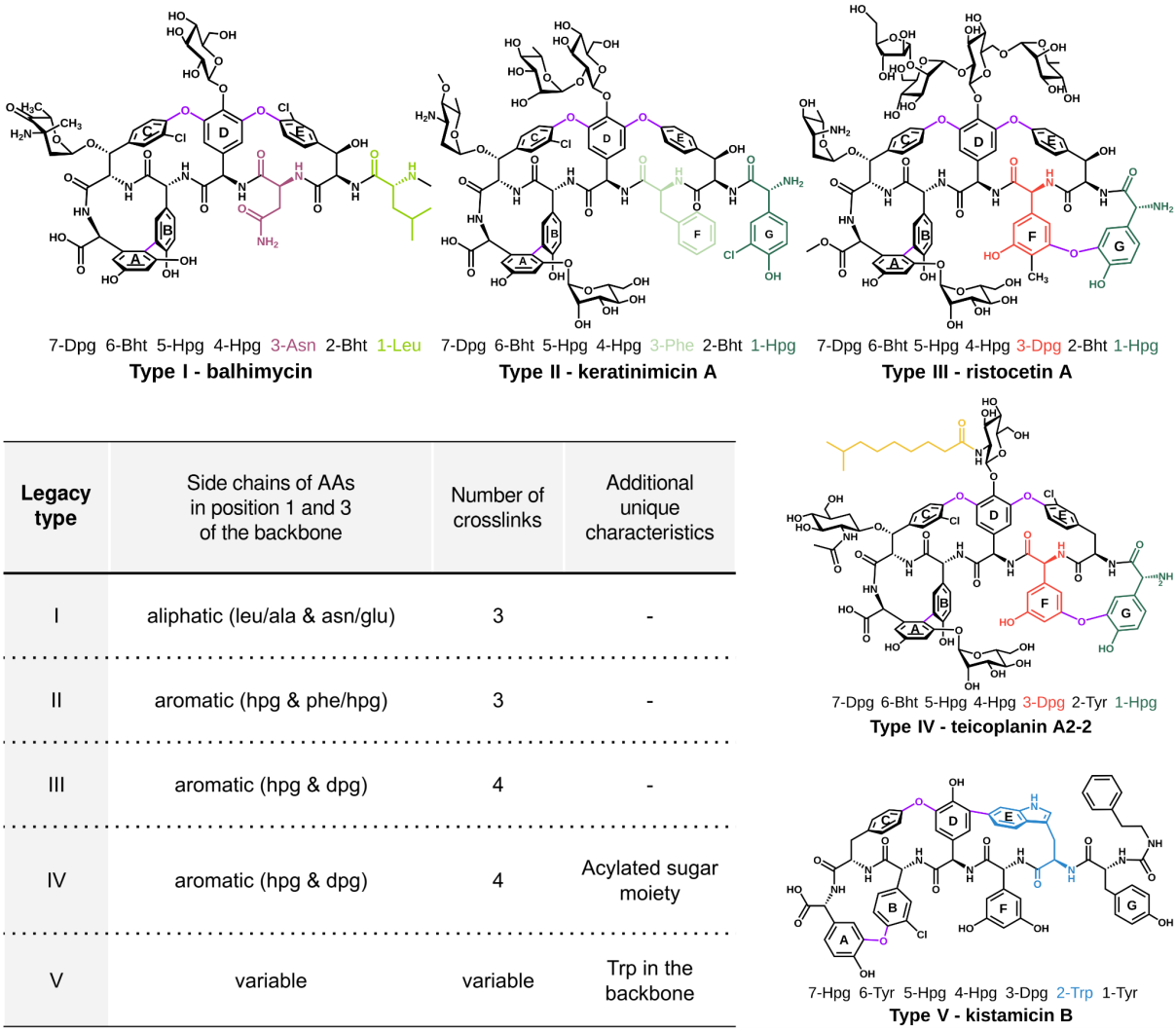
Historically, the term “glycopeptide antibiotic” has referred to a characteristic molecular architecture: a non-ribosomal peptide scaffold, typically built from a seven-amino acid backbone enriched in aromatic residues, which is conformationally constrained by oxidative crosslinks and frequently decorated by glycosylation and other tailoring reactions (e.g., halogenation, sulfation, methylation) (**Supplementary Table 1**) [1,10–13]. However, terminology around “glycopeptides” is inherently ambiguous. In broader chemical and biomedical usage, “glycopeptide” can denote any glycosylated peptide, including molecules that are biosynthetically and mechanistically unrelated to GPAs, and even structurally distinct natural products that simply carry sugar moieties. This overlap has long created confusion in the literature and across communities.

This confusion has been acknowledged before, especially for regulatory purposes in bringing new glycosylated peptides into clinical development. Accordingly, type I-IV GPAs were proposed to be named as “dalbaheptides” [14] back in 1989, a term which became the basis for naming dalbavancin a second-generation GPA [15]. However, the term dalbaheptides has not really caught on in the scientific literature. As genome mining and natural product discovery continue to expand the diversity of known glycosylated peptides, and some of them may in the future reach clinical development, the time is ripe for a revised nomenclature.

Indeed, recent discoveries have broadened the chemical and biological space of GPA-related crosslinked aromatic peptides and exposed limitations of extending the GPA label to increasingly divergent compounds. In particular, molecules previously grouped as “type V GPAs” [16,17] deviate from the classical GPA paradigm (**Figure 1**). They typically lack glycosylation, can feature variable backbone lengths (up to nine amino acids), and share distinctive structural motifs such as a tryptophane (Trp) residue crosslinked to a central 4-hydroxyphenylglycine (Hpg) [11,18]. Importantly, they also exhibit a different reported mode of action. Rather than sequestering lipid-II, they can act via inhibition of autolysins in peptidoglycan remodeling during cell division [19]. These differences raise a fundamental classification question: should these molecules be treated as an additional GPA type, or do they represent a distinct subclass within a broader class of related natural products?

To address this and building on a community consensus emerging from recent comparative work, we propose a nomenclature and classification framework that separates family-level naming from historical clinical terminology. We use “xyclopeptides” as an umbrella term for the class of crosslinked, aromatic peptide natural products and subdivide it into two subclasses: (i) the classical lipid II–binding compounds previously classified as GPA types I–IV, which we refer to as “dalabactins”, and (ii) the subclass previously termed “type V GPAs” or glycopeptide-related peptides, which we refer to as “murobactins”. The class name is intended to reflect a defining combination of chemistry and biosynthesis: xyclopeptides share a characteristic conformationally constrained (“macrocyclic-like”) peptide core generated by oxidative crosslinking and a distinctive NRPS-associated X-domain [12,20–22], a hallmark of these pathways that mediates recruitment of cytochrome P450 enzymes responsible for installing the crosslinks. This terminology aims to be clear and unambiguous, reflect shared biosynthetic specificities (NRPS assembly and oxidative crosslinking), and avoid confusion with unrelated classes of glycosylated peptides. We use dalabactins for legacy glycopeptide types I–IV to reflect their canonical target interaction, binding the D-Ala–D-Ala terminus of lipid-linked peptidoglycan precursors. We use murobactins for legacy type V to reflect their direct interaction with murein (peptidoglycan), as exemplified by members that bind peptidoglycan and inhibit cell-wall remodelling by blocking autolysin access.

To evaluate and substantiate this proposed framework, we systematically integrate analyses across chemical, genomic, and evolutionary dimensions. Specifically, we combined structural analyses of known compounds with comparative analyses of GPA-associated biosynthetic gene clusters (BGCs) and multi-locus phylogenetic reconstructions of conserved biosynthetic components. By analysing concordance and discordance between chemical similarity, BGC architecture, and evolutionary relationships, we delineate the major types within the class, identify type-specific biosynthetic signatures, and provide a practical crosswalk between legacy GPA nomenclature and the proposed xyclopeptide terminology to support consistent usage and future discovery.



Legacy type	Side chains of AAs in position 1 and 3 of the backbone	Number of crosslinks	Additional unique characteristics
I	aliphatic (leu/ala & asn/glu)	3	-
II	aromatic (hpg & phe/hpg)	3	-
III	aromatic (hpg & dpg)	4	-
IV	aromatic (hpg & dpg)	4	Acylated sugar moiety
V	variable	variable	Trp in the backbone

Figure 1: Structural characteristics of the legacy GPA types. The most important features for the classification of xyclopeptides are listed and accompanied by examples of structures. The most significant characteristics for each type are coloured. The aromatic rings of the AAs are labelled A-G based on prior publications [23]. Abbreviations: Tyr, tyrosine; Leu, leucine; Asn, asparagine; Bht, β-hydroxytyrosine; Hpg, 4-hydroxyphenylglycine; Dpg, 3,5-dihydroxyphenylglycine; Phe, phenylalanine; Trp, tryptophan. The SMILES of the structures can be found in **Supplementary Table 2**.

Results

Dataset curation of xyclopeptide-associated biosynthetic gene clusters

To compare chemical and evolutionary relationships across xyclopeptides we assembled a curated set of biosynthetic gene clusters (BGCs) associated with xyclopeptide biosynthesis. Candidate loci were first identified by automated BGC detection using antiSMASH [24] and grouped into gene cluster families to support consistent comparisons. We then manually refined BGC boundaries within each family by reviewing gene content and gene order and trimming unrelated flanking regions (see Methods: BGC dataset creation). This step was important because the antiSMASH boundary estimation can inflate apparent BGC-similarity

[25] and complicate downstream analyses, especially phylogenetic reconstructions and gene-content comparisons.

After completeness checks and boundary refinement, the final dataset comprised 182 trimmed BGCs from at least nine bacterial genera (**Supplementary Table 3**). One outlier locus from *Streptomyces varsoviensis* could not be evaluated confidently from available assemblies; therefore, the strain was resequenced and the resulting assembly was added to the dataset.

Many BGCs in the dataset are not linked to a known compound. To place these BGCs in context, we predicted likely peptide backbones from NRPS adenylation (A) domain specificity codes. Where necessary, specificity assignments were refined by comparison to established signatures. These predicted backbones were used throughout the manuscript to describe BGC diversity and to support legacy type assignments (**Supplementary Table 4**).

We labelled BGCs according to the legacy GPA types I-V, based on similarity to reference BGCs and concordant features in gene content and predicted backbone composition. In this curated dataset, BGCs corresponding to legacy types I-IV (dalabactins) and type V (murobactins) already differ in multiple features, motivating the structural comparison, BGC similarity, and phylogenetic analyses presented below.

Structure similarity analysis distinguishes dalabactins from murobactins

To assess how the known xyclopeptide structures relate to one another independently of biosynthetic gene content, we created a similarity network based on the pairwise structure similarities (**Figure 2**).

This analysis revealed a clear separation between compounds assigned to dalabactins and those assigned to murobactins. Dalabactin compounds formed a connected region in the network with similarity relationships that reflected the shared heptapeptide scaffolds and common glycosylation patterns. In contrast, murobactins clustered separately and showed limited similarity to the dalabactin cluster, consistent with the absence of glycosylation and variable peptide length. A wider spread of the murobactins across the similarity network suggests higher structural diversity.

Overall, the structure similarity network supports the view that murobactins are not simply a peripheral extension of the dalabactin structural continuum, but instead represent a distinct branch of glycopeptide-related structural diversity. This separation motivated subsequent analyses of BGC similarity and phylogeny to test whether genomic and evolutionary relationships recapitulate the same division.

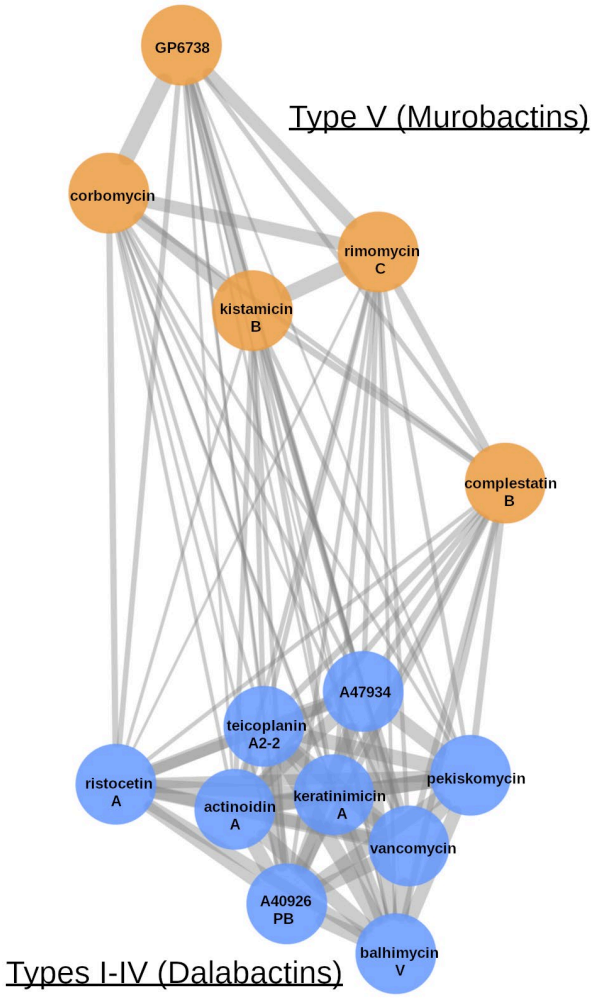


Figure 2: Structural and biosynthetic distance of murobactins. Tanimoto similarity network of the PubChem fingerprint of selected xyclopeptides structures visualized with the perfuse force-directed layout in Cytoscape [26] (see Methods, **Supplementary Table 2**). The murobactins are coloured orange, while the dalabactins are shown in blue. The width of the edges reflects the value of each similarity metric.

Phylogenetic analysis of full BGCs supports a reclassification

To test whether the structural separation between dalabactins and murobactins is mirrored by biosynthetic evolution, we carried out multi-locus phylogenetic analyses across homologous genes of our curated xyclopeptide BGCs. Because gene order is not consistently conserved across these BGCs, it cannot be used reliably for homology detection. Therefore, we first inferred fine-grained orthologous groups using zol [27] and refined ambiguous cases by phylogeny-guided curation (for example for closely related P450s; see methods section and **Supplementary Figure 1**). We then reconstructed maximum-likelihood phylogenies for individual genes based on their amino acid sequences.

Because NRPS genes are shaped by frequent domain and module rearrangements during evolution [28–33], we analysed NRPS genes at the domain level rather than as full-length genes. Phylogenies were built for each functional domain type and, where applicable, separately for each module position. This avoids comparing non-equivalent domains across modules and preserves the link between NRPS organisation and the predicted peptide backbone, including backbone length.

Gene and domain diversity was high across the dataset, and many loci were not shared broadly enough to be informative for tree building. We therefore focused on genes and domains present in at least half of the BGCs and summarised their phylogenetic signals in a super network [34] (**Figure 3**). Despite frequent incompatibilities among individual trees, a robust partition is evident: dalabactin BGCs occupy a compact region of the network, whereas murobactin BGCs span a broader phylogenetic space [35].

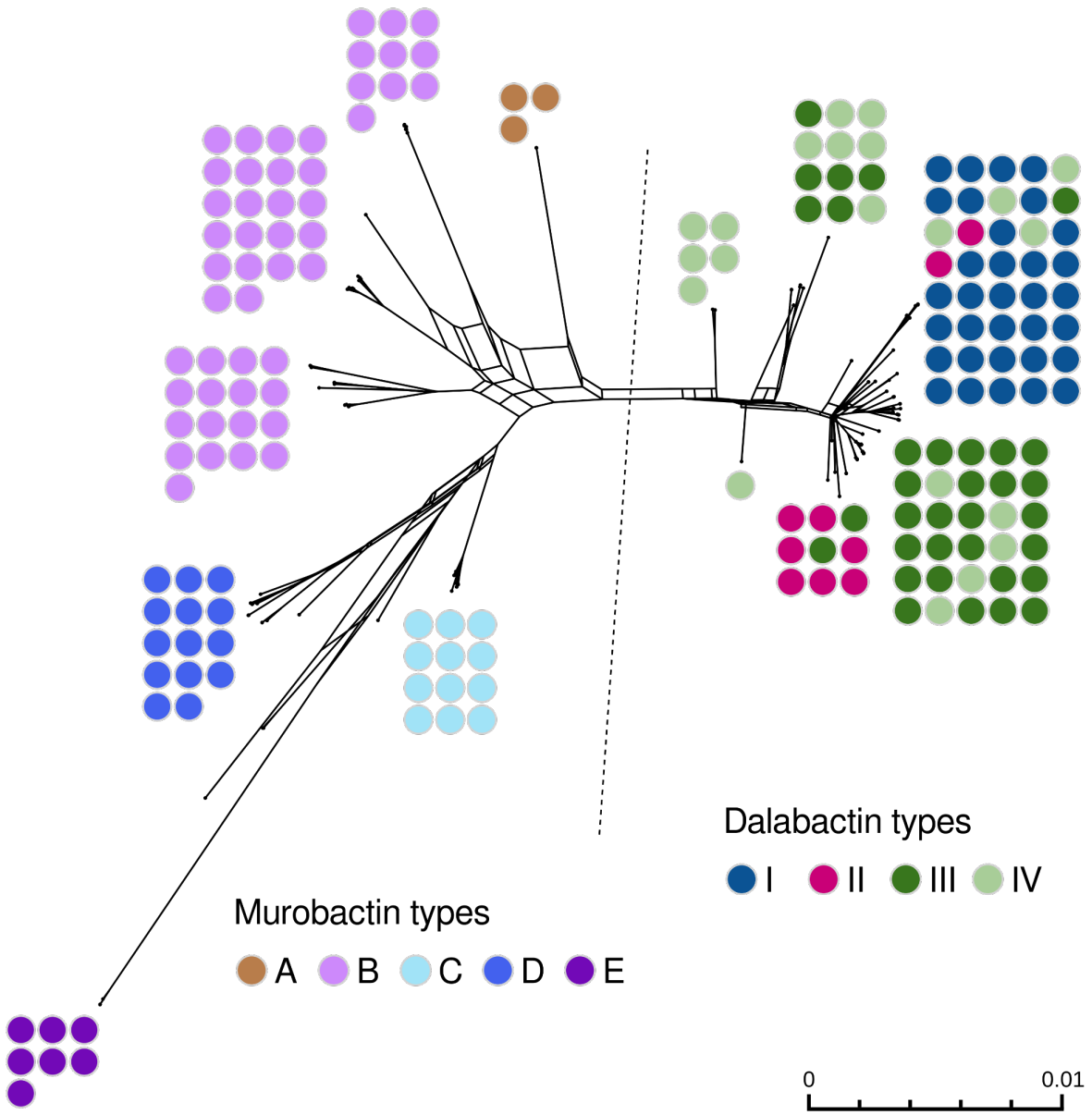


Figure 3: Phylogenetic network of xyclopeptide BGCs. Summarized super network constructed from maximum likelihood trees of all sufficiently present (>50%) genes and domains from xyclopeptide encoding BGCs (**Supplementary Table 3**) (seed=0), computed with the SplitsTree program [36], using "greedy weak compatibility" filtering to reduce visual complexity . Each node around the network corresponds to one BGC of a specific type, indicated by the colors. To simplify the visualization, close nodes are bundled in rectangular formations and a dotted line separates dalabactins and murobactins (detailed version in **Supplementary Data 1**). Most BGCs belong to one-type bundles. However, there are exceptions to this among the dalabactin BGCs, which can be explained in part by the multidimensional nature of the underlying data. Within dalabactins, some BGCs bundle closer to other types that are found in producers of the same species.

As an independent support, we also reconstructed a concatenated phylogeny using a subset of conserved genes and domains present in at least 90% of BGCs and passing a congruence filter (Methods). The concatenated tree is consistent with the super network, recovering the same deep split between dalabactins and murobactins and similar internal substructure (**Figure 4** and **Supplementary Figure 2**). We use this reference phylogeny to map gene

presence and absence patterns and to highlight subclass- and type-associated modules described in the following sections.

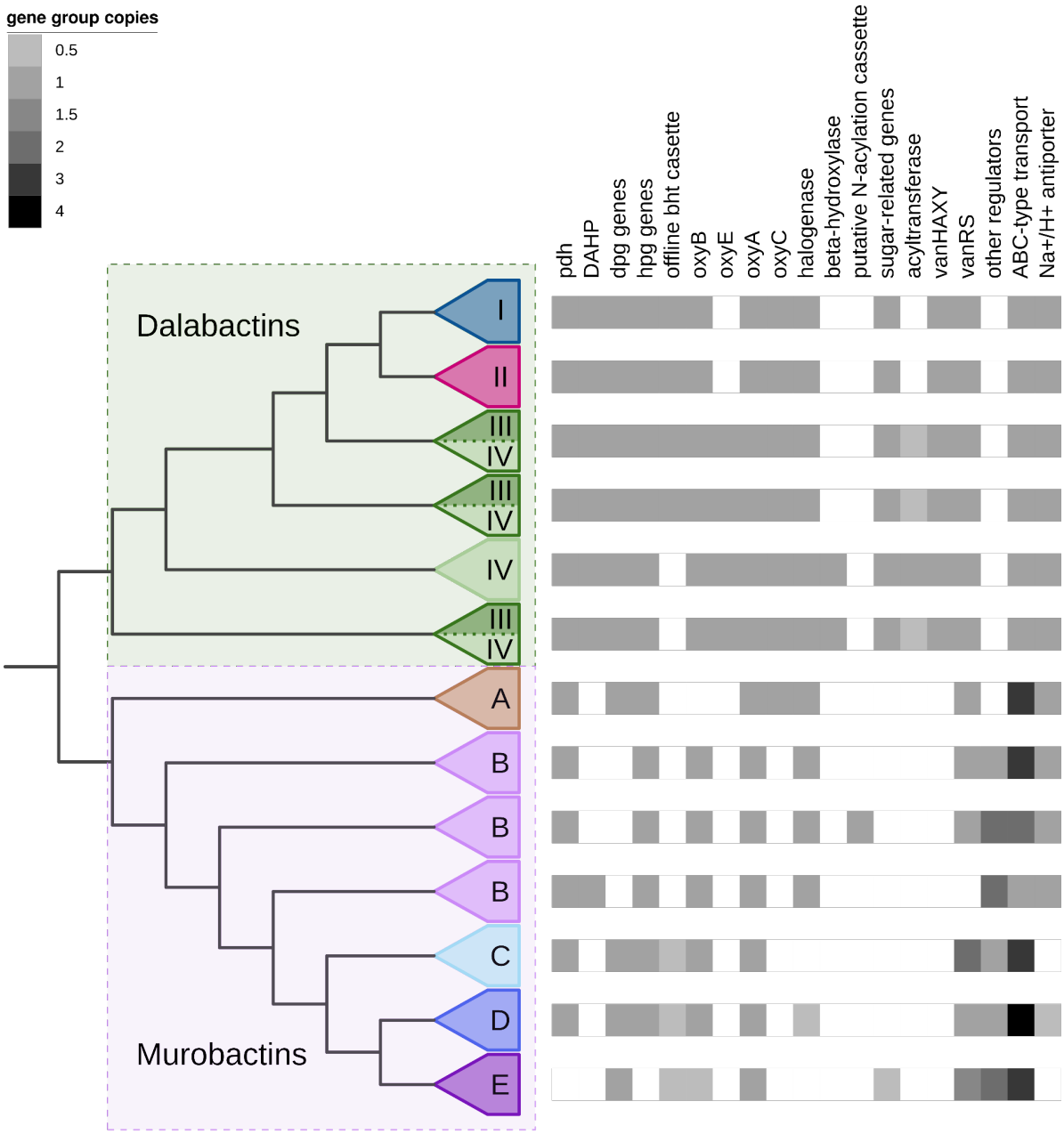


Figure 4: Phylogeny and gene patterns of xyclopeptide BGCs. Left: summarized concatenated phylogeny of xyclopeptide BGCs. Monophyletic clades were collapsed and branch lengths ignored. Colours correspond to dalabactin and murobactin types, which are labelled on the collapsed clades. Right: summarized gene presence/absense heatmap. Only genes or gene groups that form distinct patterns across xyclopeptide types are shown here. A detailed view of this figure can be seen in **Supplementary Figure 2**.

Type-specific genomic signatures within dalabactins

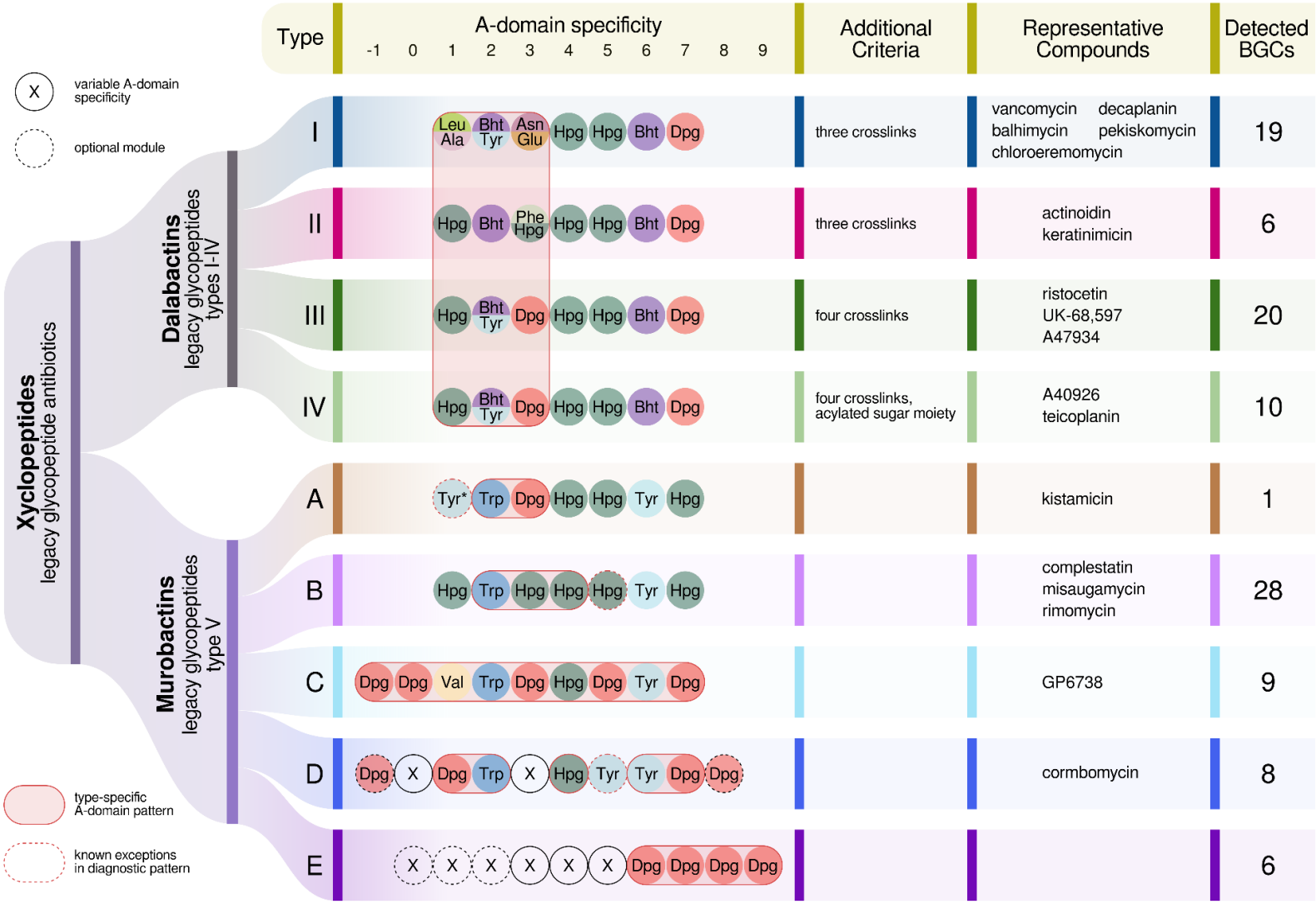

Type
A-domain specificity
-1 0 1 2 3 4 5 6 7 8 9
Additional Criteria
Representative Compounds
Detected BGCs
variable A-domain specificity
optional module
Xyclopeptides
legacy glycopeptide antibiotics
Dalabactins
legacy glycopeptides types I-IV
Murobactins
legacy glycopeptides type V
I
Leu Ala Bht Tyr Asn Glu Hpg Hpg Bht Dpg
three crosslinks
vancomycin decaplanin balhimycin pekiskomycin chloroeremomycin
19
II
Hpg Bht Phe Hpg Hpg Hpg Bht Dpg
three crosslinks
actinoidin keratinimicin
6
III
Hpg Bht Tyr Dpg Hpg Hpg Bht Dpg
four crosslinks
ristocetin UK-68,597 A47934
20
IV
Hpg Bht Tyr Dpg Hpg Hpg Bht Dpg
four crosslinks, acylated sugar moiety
A40926 teicoplanin
10
A
Tyr* Trp Dpg Hpg Hpg Tyr Hpg
kistamicin
1
B
Hpg Trp Hpg Hpg Hpg Tyr Hpg
complestatin misaugamycin rimomycin
28
C
Dpg Dpg Val Trp Dpg Hpg Dpg Tyr Dpg
GP6738
9
D
Dpg X Dpg Trp X Hpg Tyr Tyr Dpg Dpg
cormbomycin
8
E
X X X X X X Dpg Dpg Dpg Dpg
6
type-specific A-domain pattern
known exceptions in diagnostic pattern

Figure 5: Overview showing the new classification of glycopeptide antibiotics (GPAs) - now xyclopeptides - into the D-Ala–D-Ala binding subclass dalabactins (legacy GPAs type I-IV) and the murein binding subclass murobactins (legacy GPAs type V). Types: new types within each subclass stay the same for types I-IV and further distinguish legacy type V. A-domain specificity: Main distinctive criterion for the type. Red outlines indicate type specific diagnostic patterns, red dashes mark known exceptions. Positions with an "X" mark variable A-domain specificities within one type and dashed black outlines represent optional modules. The incorporated amino acids are aligned in agreement with A-domain phylogeny and numbered according to the legacy seven amino acid GPA backbone. *The first tyrosine in the structure of kistamicin (type A) does not correspond to any NRPS module in the associated BGC. Additional criteria: criteria apart from the a-domain specificity that are important to tell apart the types. Detected BGCs: number of different (< 99% sequence identity) xyclopeptide BGCs found in public sequence data (total 107).

To assess whether the established dalabactin subdivision into legacy types I–IV is supported by our expanded dataset, we compared multi-locus phylogeny, predicted backbone motifs, and gene-content patterns across curated xyclopeptide BGCs. Overall, dalabactin BGCs form a compact phylogenetic group, and legacy types are largely recovered as coherent subclades [37], with predicted backbone composition highly conserved within each type (**Figure 5**). In addition, dalabactin BGCs show the expected enrichment of glycosylation-related genes and canonical resistance genes whereas murobactin BGCs lack these genes, consistent with their established biosynthetic logic (**Supplementary Figure 2**, **Supplementary Table 1**, **Supplementary Table 5**). The main exception is the partial intermixing of legacy types III and IV, which is expected because their historical distinction is driven primarily by sugar acylation (typically a single acyltransferase) rather than broad pathway divergence. Given the overall concordance across phylogeny, backbone predictions, and key accessory modules, we retain the legacy dalabactin type I-IV classification as a practical and biologically meaningful framework.

A small number of dalabactin BGCs deviates from the typical enrichment of sugar-associated modules, including type I BGCs most similar to pekiskomycin-like loci (StsY_A, StsAN), type II BGCs related to the recently discovered kineomicin [38] loci (AkAan_1, AkAan_2, UME), and type III BGCs related to the A47934 locus (StTo, StToN1) (**Supplementary Figure 2**). Notably, apart from these rare cases and the expected III/IV boundary blur driven by sugar acylation, the legacy dalabactin types show strong concordance between chemical classification (backbone motifs and crosslink patterns) and biosynthetic organisation, reinforcing their utility for comparative analyses. In contrast, murobactins (legacy type V) span a broader phylogenetic space and show greater variability in NRPS architecture and accessory gene content, motivating a separate fine-scale classification.

Murobactins show distinct biosynthetic logic and form phylogeny-supported subclades

This divergence is also reflected in predicted backbone length and composition (**Figure 5**), supporting the separation of murobactins as a distinct subclass within xyclopeptides. In contrast to dalabactins, murobactin BGCs generally lack glycosylation-related genes and canonical resistance genes (**Supplementary Figure 2**), and instead show comparatively prominent regulatory and transport features. While tailoring capacity after backbone assembly appears more limited, several murobactin-specific genes remain unassigned. No murobactin

containing Bht has been reported, although Bht-associated genes occur sporadically in some types (**Supplementary Table 1**). Murobactin NRPSs also display expanded and distinctive domain architectures, including additional E-domains and TIGR01720 domains and, in some cases, C-starter, N-methyltransferase, and terminal inactive A-domains. Finally, multi-locus phylogeny resolves reproducible subclades associated with characteristic backbone motifs (**Figure 5**), which we use as a practical fine-scale classification into murobactin types A-E.

Murobactin type A

Type A comprises BGCs similar to kistamicin and is characterised by a Trp-Dpg motif at positions 2-3 (**Figure 5**). All identified type A BGCs encode seven NRPS modules; notably, the first module lacks an A-domain despite Tyr being present at position 1 in kistamicin. Type A most closely resembles dalabactins in overall gene content and forms an outgroup to the remaining murobactins. It uniquely includes all genes for Dpg biosynthesis and a dalabactin-like halogenase gene (**Supplementary Table 1**). Type A also encodes a distinctive set of P450 monooxygenases (Oxy) that lacks OxyB but includes a unique OxyC reported to catalyse two non-sequential crosslinks [5].

Murobactin type B

Type B murobactins also encode heptapeptide backbones but differ from type A in the absence of Dpg-specific A-domains. Their characteristic motif is Trp–Hpg–Hpg–(Hpg) at positions 2-5 (**Figure 5**), with one small exception predicted to contain Dpg in place of Hpg at position 5. Surprisingly, genes related to the precursor supply of Dpg are absent even in the two BGCs that are predicted to include Dpg in their backbone. These two BGCs are unusually small in length and lack many other characteristic genes. One structure has been recently associated with these compounds, varsomycin, where a 3,5-dihydroxyphenylalanine is present at position 5 [39]. Type B includes complestatin-, rimomycin-, and misaugamycin-associated subclades that are distinguishable phylogenetically (**Figure 4**). Regulatory repertoires vary among these subclades, and some rimomycin- and complestatin-like BGCs encode N-methyltransferase domains in module 6. Misaugamycin-like BGCs can also include a four-gene cassette proposed to support N-acylation.

Murobactin type C

Type C BGCs, represented by GP6738, are highly uniform, forming a clearly supported monophyletic group across phylogenetic representations (**Figures 3, 4**) and exhibiting consistent gene-content patterns (**Figure 4, Supplementary Figure 2**). Predicted backbones comprise nine residues and include a Val residue at position 3 followed by the conserved Trp motif typical of murobactins (**Figure 5**). Type C BGCs encode precursor supply genes for Dpg and Hpg and include an alpha/beta hydrolase associated with offline Bht hydroxylation, but lack the remaining genes required for Bht biosynthesis; accordingly, Bht-specific A-domain signatures are not detected and the known representative structure does not contain Bht. Type C BGCs encode only two *Oxy* genes, consistent with fewer crosslinks, and are unique among murobactins in encoding two copies of the VanR/VanS two-component system.

Murobactin type D

Type D encompasses backbones ranging from eight to ten residues and includes corbomycin-like BGCs (**Figure 5**). Although more variable than types A-C, type D displays recurring motifs (**Supplementary Figure 3**), including a conserved Dpg-Trp motif at aligned positions 1-2 and a conserved Hpg at aligned position 4, which is shared across types A-D and is likely linked

to the conserved Trp with a crosslink. A Tyr-Dpg motif at aligned positions 6-7 is also common. With respect to precursor supply, type D resembles type C, but it shows additional conserved regulatory and transport features, including a LysR-like regulator and two ABC transporter classes. A halogenase occurs in a subclade but is not universally conserved. Several recurrent genes remain unassigned and may encode type-specific functions.

Murobactin type E

Type E is the most heterogeneous murobactin group and currently lacks associated structures. Predicted backbones range from seven to ten residues and are characterised by a Dpg-rich tetrapeptide tail (**Supplementary Figure 4**), while typically lacking Trp and Hpg. Type E BGCs consistently lack shikimate-pathway genes (**Supplementary Table 1**) and show substantial variation in NRPS domain organisation, mirroring backbone diversity. It is the only type to display two distinct oxy-gene patterns. A small subclade uniquely encodes a mannosyltransferase, and one “minimal” BGC resembles the compact architectures noted in type B, lacking several transport and regulatory genes common in other xyclopeptide clusters. Given the absence of experimentally characterised structures and limited functional data, type E should be considered provisional and is likely to be refined as additional murobactins are discovered and characterised.

Discussion

The historical type I-V classification for glycopeptide antibiotics was built around chemical features and was invaluable for organising an emerging field. However, accumulating evidence shows that legacy type V BGCs and their corresponding compounds are consistently separated from legacy types I–IV in terms of structure, biosynthetic organization, and mechanism of action that targets peptidoglycan remodelling rather than binding lipid II. This has led to recurring confusion in the literature, including the fact that many legacy type V molecules are not glycosylated, and therefore do not fit the term “glycopeptide antibiotic” in a strict sense. Furthermore, the interpretation of glycopeptides as any glycosylated peptide does not reflect the 7-10 AA backbone with characteristic crosslinks and antibiotic activity. Against this backdrop, a nomenclature update is not cosmetic, it is necessary to align names with biology and to prevent continued overloading of “type V” as a catch-all bin for newly discovered outliers.

We therefore propose xyclopeptides as an umbrella framework that preserves relatedness while making the major split explicit. Within xyclopeptides, dalabactins correspond to legacy GPA types I-IV, and murobactins correspond to legacy type V. These names were chosen deliberately to be (i) distinctive and unambiguous, (ii) expandable as additional chemistry and BGC diversity is discovered, and (iii) compatible with legacy usage by retaining type labels where they remain informative. Importantly, the framework is intended to be practical for comparative genomics and discovery, it groups what is genuinely related while separating subclasses that differ in target biology and biosynthetic logic.

There have been other attempts to reclassify type I-IV GPAs in the past, such as the suggested “dalbaheptides”[14], which was the basis for naming the second-generation GPA dalbavancin [ref]. However, the term dalbaheptides has not really caught on in the scientific literature.

A central goal of this study was to test whether this nomenclature is supported beyond intuition from known structures and mechanisms. Across multiple, independent analyses, the dalabactin–murobactin split is consistently recovered from BGC evolution, and murobactins span a broader phylogenetic space than dalabactins, mirroring their broader variability in NRPS architecture and predicted backbone patterns. These evolutionary results align with chemical and functional distinctions reported for representative murobactins such as complestatin, corbomycin, and GP6738, reinforcing that the proposed names capture real biological structure rather than arbitrary clustering. At finer scale, the reproducible subclades within murobactins provide a pragmatic basis for the provisional A-E subdivision, which can be refined and expanded as additional structures and modes of action are characterised.

At the same time, we emphasise limitations. BGC boundary definition and orthology inference across diverse BGCs remain non-trivial, and reticulate evolution, including horizontal transfer [13,31,40], complicates any single “true” tree. Our approach mitigates this by combining locus-wise phylogenies, network representations that tolerate phylogenetic incongruence, and a conservative concatenation strategy. But ambiguous cases remain, particularly for murobactin type E, where no experimentally linked structures are currently available. The A-E scheme should therefore be viewed as a working classification that is designed to be updated, not a final taxonomy.

Finally, we want to stress that durable nomenclature changes require broad consensus. In developing this framework, we benefited from extensive discussions among researchers studying chemistry, biosynthesis, genomics, and mechanism of action. This collaborative approach matters because naming systems succeed when they encode shared agreement and remain straightforward to apply, including for non-specialists. A useful model comes from the RiPP field, where community-driven nomenclature and classification helped stabilise terminology across a rapidly expanding chemical space [41]. We hope xyclopeptides can play a similar role for dalabactins and murobactins by reducing ambiguity, strengthening links between chemical and biosynthetic classes, and providing a scalable structure for future discoveries. More broadly, we hope this process can serve as a template for future nomenclature efforts in other natural product families, and as an example of how collaborative, cross-disciplinary discussion can improve the long-term clarity and usability of scientific language.

Methods

BGC dataset creation

For the evolutionary analysis of xyclopeptide BGCs to be as complete as possible, we aimed to create a dataset of all sequenced clusters. The initial dataset contained all known clusters found in the literature [9,11,19,42–62] and was extended by searching in publicly available sequence databases (MIBiG [63], most of NCBI RefSeq and Genbank [64–66], JGI IMG DB [67], MGnify [68]) and some published projects [69–72]. Information on the origin of the BGCs, the date of the search and more metadata are available in **Supplementary Table 3**. **Supplementary Data 4** includes the accession numbers of all NCBI (RefSeq and Genbank)

entries checked and whether they were a hit (True), not a hit (False) and the ones where the search failed due to lack of protein sequences associated with the accession number (Failed). The target of the search, accommodated via HMMER (v. 3.3.1, accessed from http://hmmer.org/, RRID:SCR_005305), was the X-domain in the last NRPS module (its HMM was extracted from antiSMASH v6 [73], RRID:SCR_022060), which has so far only been found in xyclopeptide encoding BGCs [72,73]. However, the X-domain has evolved from a condensation domain (C domain) and their sequences remain quite similar [6]. To avoid false positives, the known clusters were searched with the X-domain HMM and the lowest local score of the confirmed X-domains (300) was used as a minimal threshold. Lower local scores (up to 250) were manually checked for the MIBiG dataset and indeed no X-domains were found below the chosen threshold. The custom scripts used for this and all other parts of the analysis can be found in **Supplementary Data 5**.
The sequences that were a hit in the search were used as input for an antiSMASH v7 [24] (RRID:SCR_022060) analysis (default parameters + MIBiG cluster comparison) to detect the full BGC. At this stage, after manual inspection, some candidates were dropped due to low quality assemblies which led to obviously incomplete clusters, and due to a few false positives (where a C domain was mistakenly detected as an X-domain). The cluster regions that passed this inspection were assigned an ID reflecting their taxonomic placement (**Supplementary Table 3**) and their coding domains were re-annotated with bakta (v1.8.1) [74] to ensure homogeneity. Bakta was run with default parameters + skipping trna and tmrna detection (--skip-trna --skip-tmrna) and on metagenome mode when appropriate. The original contig headers were kept (--keep-contig-headers) but the assigned ID was used as a locus and locus tag prefix. The bakta-annotated regions were re-analyzed with antiSMASH v7 [24] with all features on (--fullhmmer –clusterhmmer –tigrfam –asf –cc-mibig –cb-general –cb-subclusters –cb-knownclusters –pfam2go –rre –smcog-trees –tfbs).
One exception to this process was done for a very unusual BGC from *Streptomyces varsoviensis*, which could not be conclusively labelled complete. Due to its interesting characteristics though, the genome of the strain was resequenced. For isolation of high molecular weight genomic DNA *S. varsoviensis* was cultured in R5 medium [75] for 2 days on a rotary shaker at 28 °C. DNA isolation was performed by using DNA isolation kit (NucleoBond® HMW DNA Kit, Machery-Nagel, Düren, Germany) following the manufacturer's protocol. Genome sequencing was performed by the NGS competence center (NCC in Tübingen, Germany) on a PromethION (Nanopore) with 9.4.1 chemistry (details in **Supplementary Data 6**). Assembly was done with Unicycler [76] (for long reads) and corrected with medaka (accessed from https://github.com/nanoporetech/medaka) based on the sequencing parameters (**Supplementary Data 6**). The resulting genome sequence was searched for the X-domain as described above and BGCs were included in the analysis (**Supplementary Table 3**). The sequence was also submitted to NCBI and was assigned the accession number CP162614.
To overcome antiSMASH's generous selection of BGC boundaries, a manual inspection was necessary, to ensure the quality of the evolutionary reconstruction, which would have been affected by the accidental inclusion of unrelated sequences. The process was sped up by annotating the BGCs in groups of GCFs as defined by a BiG-SCAPE analysis [77] (v1.0.1 2020-01-27, RRID:SCR_022561). The chosen threshold of 0.2 defined the largest possible groups that never have more than one type xyclopeptide (based on the MIBiG dataset). BiG-SCAPE was run on glocal mode, mixing all classes (as some BGCs are marked NRPS and others are marked other), but using the score weights of the NRPS class. The GCFs were then visualized with clinker [78] (v0.0.28). We meticulously manually curated each cluster by

investigating every single gene for its function and possible role in the biosynthesis and the clinker visualization ensured uniform trimming of the most closely related BGCs (**Supplementary Figure 5**). Thus, the final dataset of 182 trimmed clusters from 9 different genera was created (**Supplementary Table 3**).
To put a number on how many BGCs per type we have in our dataset (**Figure 5**), we clustered the 182 trimmed BGCs on DNA level with MMseqs2 (version 18.8cc5c, RRID:SCR_022962) in the easy-cluster-mode with a minimal sequence identity and coverage of 99% (--min-seq-id 0.99 -c 0.99 --cov-mode 0), which resulted in 107 BGCs.

Chemical structure similarity network

The structures of the glycopeptides (**Figure 1, Figure 2**) were collected from the original publications [11,17,19,42,43,50,51,53,79–88] (entries under database: MIBiG in **Supplementary Table 3**), formatted and converted to SMILES with ChemDraw (https://revvitysignals.com/products/research/chemdraw, RRID:SCR_016768) and RDKit (v. 2020.09.1, http://www.rdkit.org/, RRID:SCR_014274).
The examination of the structural variation within and among xyclopeptide types was done by use of the Tanimoto similarity metric [89]. For the PubChem fingerprint, the SMILES of known GPA compounds were inserted into the Cytoscape program (v. 3.9.1, RRID:SCR_003032) [26], which accommodates the calculation and visualisation of a structure similarity network (**Figure 2**) based on Tanimoto with *chemViz2* (fingerprint: Pubchem), a cheminformatics app for Cytoscape. The RDKit, Morgan [90] and MACCS 166 [91] fingerprints were calculated using RDKit with the default settings for the RDKit and MACCS fingerprint and a radius of two for the Morgan fingerprint. The Tanimoto similarity values were imported in Cytoscape and visualised in **Supplementary Figure 7**. The nodes of the resulting network were annotated by legacy GPA type (I-IV or V) and the edge width was adjusted to represent the degree of similarity among the connecting parts.

A-domain specificity code analysis

A-domain specificity was predicted (**Supplementary Table 4**) using the 34 AA long specificity-conferring code determined by NRPSpredictor2, implemented in antiSMASH (v. 5.1.2, RRID:SCR_022060) [92–94]. Due to the fact that certain A-domain specificities could not be predicted *in silico* [42], all the known amino acid specificities (for glycopeptides whose structure is known) were used in a blastp (v2.14.0+) search [95] and best scoring alignments were used to determine the final annotations, *i.e.*, the ones used for the rest of the analysis, which sometimes differed from the antiSMASH assigned ones. A summary of A domain specificity was visualised in **Figure 5**, while the predicted backbone of all BGCs can be seen in **Supplementary Figure 2.**

Detection of homologous genes

The identification of homologous groups of genes from the full dataset (**Supplementary Table 5**) was mostly carried out by the zol tool (v 1.3.9) [27], which infers phylogenetic orthology for comparative genomics of gene clusters. The platform (run with default parameters) generated 309 so-called orthogroups (OGs). There was one case of related genes with distinct functions being placed in the same Orthogroup, namely the oxyAs and oxyEs, which are known to be

closely related [5,43]. Those proteins were used for a multiple sequence alignment (MSA) and their separation into groups was guided by their phylogenetic placement compared to proteins of known function. Cases of OGs with only a few clusters containing multiple copies of a gene were dealt with via a custom script, discarding some copies based on similarity criteria [96]. Finally, the NRPS domains were extracted from the genes and split into groups based on the position of their module compared to the rest.
The latter was achieved by building an MSA from the concatenated 34 AA A-domain specificity codes (as provided in the antiSMASH results [24,97] with the help of the NRPSpredictor2 tool [92]) of the underlying A-domains (**Supplementary Data 3**). The longest conserved region (positions 4-7 of balhimycin) was taken as a center and all internal gaps were rejected, since the positions directly correspond to the backbone of the compounds. The alignment was in agreement with the fact that positions 1-3 of the types I-IV glycopeptides are known to be the most variable among known compounds and their corresponding BGCs [37]. All module positions (00, 0, 1-9) were annotated based on this fixed MSA, with the balhimycin encoding BGC as a template for positions 1-7.

Visualising complete evolutionary history

All multiple sequence alignments (MSAs) were performed with the mafft tool (v7.490, 2021/Oct/30, RRID:SCR_011811) with default parameters [98] (**Supplementary Data 7**). Phylogenetic trees for every occasion in this study (153 gene and domain trees, as well as concatenated phylogenies described below) were built with iqtree (multicore v2.2.0.3 COVID-edition for Linux 64-bit built Aug 2 2022) [99,100] (**Supplementary Data 7**). For the gene and domain trees, iqtree was first run only in model testing mode, checking for all bacteria-suitable evolutionary models (suitable models are listed in **Supplementary table 6**) and then in tree-building mode based on the best fitting model. The resulting trees can be seen in **Supplementary Data 1**.
A graphical summary of the adequately populated (n>50%) 47 separate (partial) gene and domain trees was calculated and visualised by the super network algorithm [101] (default options) implemented in the SplitsTree CE tool (version 6.0.10-beta) [36,102,103] (**Figure 3**) after greedily selecting a weakly compatible set of splits of maximum support (GreedyWeaklyCompatible splits filter). The super network (**Supplementary Data 1**) summarises the set of input trees, taking into account that many of the trees are incomplete. It is a splits network in which each band of parallel edges represents one of the splits or branches found in the set of input trees. Incompatibilities among the input trees give rise to parallelograms in the network. Edges in the network are scaled to represent the average relative length of the corresponding edges in the input trees. The set of splits computed by the super network method was greedily filtered by decreasing support (number of trees that contain a given split), so as to obtain a subset of “weakly compatible splits” that maintains major incompatibilities, while avoiding higher-dimensional edge configurations in the network, and thereby visual clutter.

Concatenated phylogeny

A species phylogeny can be constructed from concatenated sequences of core genes, as long as their separate trees are congruent [104]. Following this concept, 22 genes or domains found in at least 90% of the BGCs in the dataset were identified. Duplications were considered

acceptable up to a max of 5% (i.e. up to 5% of the clusters had more than one gene in the orthologous group) but otherwise, these groups were single-copy genes. These two thresholds were chosen based on a wide-scale phylogenetic study [104]. However, it was necessary to ensure the congruence of the participating genes before concatenating them. The underlying MSAs were first trimmed with trimAl (v1.4.rev15 build[2013-12-17], RRID:SCR_017334) [105] with default parameters and then the trees were recalculated (**Supplementary Data 7**). The latter were used in a congruence analysis as performed by Parks and colleagues [104]: The well supported splits were calculated by checking their existence in random subsampled concatenated phylogenies and then for each gene tree, their presence was used to calculate 'normalised compatible split length' ,a metric that reflects congruence of this tree to the rest in the group (**Supplementary Table 7**). This value was computed with a python script, implemented with Biopython [106] (RRID:SCR_007173), FastTree v2.1.11 [107] (RRID:SCR_015501) and a function of the (still in development) GeneTreeTk toolbox (accessed from https://github.com/dparks1134/GeneTreeTk). Eleven genes or domains passed a specific threshold (0.67) and were used for the concatenated phylogeny representing the evolutionary history of the xyclopeptide BGCs (**Supplementary Data 2**). The tree was then rooted with the MAD algorithm [108]. This tree, together with an absence/presence heatmap of the various genes and domains present in the clusters was visualised with iTOL [109,110] (RRID:SCR_018174) (**Supplementary Figure 2**). The tree in **Figure 4** was rooted between the dalabactins and murobactins. This alternative rooting is quite plausible due to the topology of the super network (**Figure 3**), supported also by the long branch leading to type A murobactins in the concatenated phylogeny (**Supplementary Figure 2**).

References

1. Stegmann E, Frasch HJ, Wohlleben W. Glycopeptide biosynthesis in the context of basic cellular functions. Current Opinion in Microbiology. 2010. pp. 595–602. doi:10.1016/j.mib.2010.08.011

2. Butler MS, Hansford KA, Blaskovich MAT, Halai R, Cooper MA. Glycopeptide antibiotics: back to the future. J Antibiot (Tokyo). 2014;67: 631–644.

3. Higgins HM, Harrison WH, Wild GM, Bungay HR, McCORMICK MH. Vancomycin, a new antibiotic. VI. Purification and properties of vancomycin. Antibiot Annu. 1957;5: 906–914.

4. Hansen MH, Stegmann E, Cryle MJ. Beyond vancomycin: recent advances in the modification, reengineering, production and discovery of improved glycopeptide antibiotics to tackle multidrug-resistant bacteria. Curr Opin Biotechnol. 2022;77: 102767.

5. Greule A, Izoré T, Iftime D, Tailhades J, Schoppet M, Zhao Y, et al. Kistamicin biosynthesis reveals the biosynthetic requirements for production of highly crosslinked glycopeptide antibiotics. Nat Commun. 2019;10: 2613.

6. Schoppet M, Peschke M, Kirchberg A, Wiebach V, Süssmuth RD, Stegmann E, et al. The biosynthetic implications of late-stage condensation domain selectivity during glycopeptide antibiotic biosynthesis. Chem Sci. 2019;10: 118–133.

7. Kittilä T, Kittel C, Tailhades J, Butz D, Schoppet M, Büttner A, et al. Halogenation of glycopeptide antibiotics occurs at the amino acid level during non-ribosomal peptide synthesis. Chem Sci. 2017;8: 5992–6004.

8. Wohlleben W, Stegmann E, Süssmuth RD. Chapter 18. Molecular genetic approaches to analyze glycopeptide biosynthesis. Methods Enzymol. 2009;458: 459–486.

9. Truman AW, Kwun MJ, Cheng J, Yang SH, Suh J-W, Hong H-J. Antibiotic resistance mechanisms inform discovery: identification and characterization of a novel amycolatopsis strain producing ristocetin. Antimicrob Agents Chemother. 2014;58: 5687–5695.

10. Chen S, Wu Q, Shen Q, Wang H. Progress in Understanding the Genetic Information and Biosynthetic Pathways behind Amycolatopsis Antibiotics, with Implications for the Continued Discovery of Novel Drugs. Chembiochem. 2016;17: 119–128.

11. Xu M, Wang W, Waglechner N, Culp EJ, Guitor AK, Wright GD. GPAHex-A synthetic biology platform for Type IV-V glycopeptide antibiotic production and discovery. Nat Commun. 2020;11: 5232.

12. Yim G, Thaker MN, Koteva K, Wright G. Glycopeptide antibiotic biosynthesis. Journal of Antibiotics. 2014. pp. 31–41. doi:10.1038/ja.2013.117

13. Donadio S, Sosio M, Stegmann E, Weber T, Wohlleben W. Comparative analysis and insights into the evolution of gene clusters for glycopeptide antibiotic biosynthesis. Mol Genet Genomics. 2005;274: 40–50.

14. Parenti F, Cavalleri B. Proposal to name the vancomycin-ristocetin like glycopeptides as dalbaheptides. J Antibiot (Tokyo). 1989;42: 1882–1883.

15. Olsufyeva EN. History of the Creation of a New Generation of Antibiotics of the Group of Polycyclic Glycopeptides. Molecular Biology. 2024;58: 980–1002.

16. Seto H, Fujioka T, Furihata K, Kaneko I, Takahashi S. Structure of complestatin, a very strong inhibitor of protease activity of complement in the human complement system. Tetrahedron Lett. 1989;30: 4987–4990.

17. Naruse N, Oka M, Konishi M, Oki T. New antiviral antibiotics, kistamicins A and B. II. Structure determination. J Antibiot (Tokyo). 1993;46: 1812–1818.

18. Tian L, Shi S, Zhang X, Han F, Dong H. Newest perspectives of glycopeptide antibiotics: biosynthetic cascades, novel derivatives, and new appealing antimicrobial applications. World J Microbiol Biotechnol. 2023;39: 67.

19. Culp EJ, Waglechner N, Wang W, Fiebig-Comyn AA, Hsu Y-P, Koteva K, et al. Evolution-guided discovery of antibiotics that inhibit peptidoglycan remodelling. Nature. 2020. doi:10.1038/s41586-020-1990-9

20. Miller BR, Gulick AM. Structural Biology of Nonribosomal Peptide Synthetases. Methods Mol Biol. 2016;1401: 3–29.

21. Izoré T, Candace Ho YT, Kaczmarski JA, Gavriilidou A, Chow KH, Steer DL, et al. Structures of a non-ribosomal peptide synthetase condensation domain suggest the basis of substrate selectivity. Nat Commun. 2021;12: 2511.

22. Luo L, Kohli RM, Onishi M, Linne U, Marahiel MA, Walsh CT. Timing of epimerization and condensation reactions in nonribosomal peptide assembly lines: kinetic analysis of

phenylalanine activating elongation modules of tyrocidine synthetase B. Biochemistry. 2002;41: 9184–9196.

23. Nicolaou KC, Boddy CN, Bräse S, Winssinger N. Chemistry, Biology, and Medicine of the Glycopeptide Antibiotics. Angew Chem Int Ed Engl. 1999;38: 2096–2152.

24. Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, et al. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res. 2023;51: W46–W50.

25. Chavali AK, Rhee SY. Bioinformatics tools for the identification of gene clusters that biosynthesize specialized metabolites. Brief Bioinform. 2018;19: 1022–1034.

26. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13: 2498–2504.

27. Salamzade R, Tran PQ, Martin C, Manson AL, Gilmore MS, Earl AM, et al. zol and fai: large-scale targeted detection and evolutionary investigation of gene clusters. Nucleic Acids Res. 2025;53. doi:10.1093/nar/gkaf045

28. Chevrette MG, Gavrilidou A, Mantri S, Selem-Mojica N, Ziemert N, Barona-Gómez F. The confluence of big data and evolutionary genome mining for the discovery of natural products. Nat Prod Rep. 2021;38: 2024–2040.

29. Waglechner N, McArthur AG, Wright GD. Phylogenetic reconciliation reveals the natural history of glycopeptide antibiotic biosynthesis and resistance. Nature Microbiology. 2019;4: 1862–1871.

30. Baunach M, Chowdhury S, Stallforth P, Dittmann E. The Landscape of Recombination Events That Create Nonribosomal Peptide Diversity. Mol Biol Evol. 2021;38: 2116–2130.

31. Medema MH, Cimermancic P, Sali A, Takano E, Fischbach MA. A Systematic Computational Analysis of Biosynthetic Gene Cluster Evolution: Lessons for Engineering Biosynthesis. PLoS Comput Biol. 2014;10. doi:10.1371/journal.pcbi.1004016

32. Jenke-Kodama H, Dittmann E. Bioinformatic perspectives on NRPS/PKS megasynthases: Advances and challenges. Natural Product Reports. Royal Society of Chemistry; 2009. pp. 874–883. doi:10.1039/b810283j

33. Ziemert N, Podell S, Penn K, Badger JH, Allen E, Jensen PR. The natural product domain seeker NaPDoS: A phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. PLoS One. 2012;7. doi:10.1371/journal.pone.0034064

34. Huson DH, Rupp R, Scornavacca C. Phylogenetic networks from trees. Phylogenetic Networks. Cambridge: Cambridge University Press; 2010. pp. 265–299.

35. Zeng D, Debabov D, Hartsell TL, Cano RJ, Adams S, Schuyler JA, et al. Approved Glycopeptide Antibacterial Drugs: Mechanism of Action and Resistance. Cold Spring Harb Perspect Med. 2016;6. doi:10.1101/cshperspect.a026989

36. Huson DH, Bryant D. The SplitsTree App: interactive analysis and visualization using phylogenetic trees and networks. Nat Methods. 2024;21: 1773–1774.

37. Hansen MH, Adamek M, Iftime D, Petras D, Schuseil F, Grond S, et al. Resurrecting

ancestral antibiotics: unveiling the origins of modern lipid II targeting glycopeptides. Nat Commun. 2023;14: 7842.

38. Yushchuk O, Berini F, Zhong L, Rückert-Reed C, Bernasconi E, Bartolone L, et al. A rare peptide scaffold in kineomicins, the glycopeptide antibiotics produced by Actinokineospora auranticolor DSM 44650. Communications Chemistry. 2025;8: 134.

39. Li X, Wang M, Yin L, Cao J, Wang R, Guo W, et al. Phylogeny-Informed Discovery of New Glycopeptide Antibiotics Produced by a Minimal Biosynthetic Gene Cluster. Org Lett. 2025;27: 2444–2449.

40. Adamek M, Alanjary M, Sales-Ortells H, Goodfellow M, Bull AT, Winkler A, et al. Comparative genomics reveals phylogenetic distribution patterns of secondary metabolites in Amycolatopsis species. BMC Genomics. 2018;19: 1–15.

41. Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat Prod Rep. 2013;30: 108–160.

42. Thaker MN, Wang W, Spanogiannopoulos P, Waglechner N, King AM, Medina R, et al. Identifying producers of antibacterial compounds by screening for antibiotic resistance. Nat Biotechnol. 2013;31: 922–927.

43. Xu M, Wang W, Waglechner N, Culp EJ, Guitor AK, Wright GD. Phylogeny-Informed Synthetic Biology Reveals Unprecedented Structural Novelty in Type V Glycopeptide Antibiotics. ACS Cent Sci. 2022;8: 615–626.

44. Pootoolal J, Thomas MG, Marshall CG, Neu JM, Hubbard BK, Walsh CT, et al. Assembling the glycopeptide antibiotic scaffold: The biosynthesis of A47934 from Streptomyces toyocaensis NRRL15009. Proc Natl Acad Sci U S A. 2002;99: 8962–8967.

45. Pelzer S, Süssmuth R, Heckmann D, Recktenwald J, Huber P, Jung G, et al. Identification and analysis of the balhimycin biosynthetic gene cluster and its use for manipulating glycopeptide biosynthesis in Amycolatopsis mediterranei DSM5908. Antimicrob Agents Chemother. 1999;43: 1565–1573.

46. Wink JM, Kroppenstedt RM, Ganguli BN, Nadkarni SR, Schumann P, Seibert G, et al. Three new antibiotic producing species of the genus Amycolatopsis, Amycolatopsis balhimycina sp. nov., A. tolypomycina sp. nov., A. vancoresmycina sp. nov., and description of Amycolatopsis keratiniphila subsp. keratiniphila subsp. nov. and A. keratiniphila subsp. nogabecina subsp. nov. Syst Appl Microbiol. 2003;26: 38–46.

47. Xu L, Huang H, Wei W, Zhong Y, Tang B, Yuan H, et al. Complete genome sequence and comparative genomic analyses of the vancomycin-producing Amycolatopsis orientalis. BMC Genomics. 2014;15: 363.

48. Labeda DP. Amycolatopsis coloradensis sp. nov., the Avoparcin (LL-AV290)-Producing Strain. Int J Syst Bacteriol. 1995;45: 124–127.

49. Wink J, Gandhi J, Kroppenstedt RM, Seibert G, Sträubler B, Schumann P, et al. Amycolatopsis decaplanina sp. nov., a novel member of the genus with unusual morphology. Int J Syst Evol Microbiol. 2004;54: 235–239.

50. Xu F, Wu Y, Zhang C, Davis KM, Moon K, Bushin LB, et al. A genetics-free method for high-throughput discovery of cryptic microbial metabolites. Nat Chem Biol. 2019;15:

161–168.

51. Yim G, Kalan L, Koteva K, Thaker MN, Waglechner N, Tang I, et al. Harnessing the synthetic capabilities of glycopeptide antibiotic tailoring enzymes: characterization of the UK-68,597 biosynthetic cluster. Chembiochem. 2014;15: 2613–2623.

52. Sosio M, Kloosterman H, Bianchi A, de Vreugd P, Dijkhuizen L, Donadio S. Organization of the teicoplanin gene cluster in Actinoplanes teichomyceticus. Microbiology (Reading). 2004;150: 95–102.

53. Yushchuk O, Vior NM, Andreo-Vidal A, Berini F, Rückert C, Busche T, et al. Genomic-Led Discovery of a Novel Glycopeptide Antibiotic by DSM 45129. ACS Chem Biol. 2021;16: 915–928.

54. Sosio M, Stinchi S, Beltrametti F, Lazzarini A, Donadio S. The gene cluster for the biosynthesis of the glycopeptide antibiotic A40926 by nonomuraea species. Chem Biol. 2003;10: 541–549.

55. Bardone MR, Paternoster M, Coronelli C. Teichomycins, new antibiotics from Actinoplanes teichomyceticus nov. sp. II. Extraction and chemical characterization. J Antibiot (Tokyo). 1978;31: 170–177.

56. Chiu HT, Hubbard BK, Shah AN, Eide J, Fredenburg RA, Walsh CT, et al. Molecular cloning and sequence analysis of the complestatin biosynthetic gene cluster. Proc Natl Acad Sci U S A. 2001;98: 8548–8553.

57. Kwun MJ, Hong H-J. Genome Sequence of Streptomyces toyocaensis NRRL 15009, Producer of the Glycopeptide Antibiotic A47934. Genome Announc. 2014;2. doi:10.1128/genomeA.00749-14

58. Banik JJ, Craig JW, Calle PY, Brady SF. Tailoring enzyme-rich environmental DNA clones: a source of enzymes for generating libraries of unnatural natural products. J Am Chem Soc. 2010;132: 15661–15670.

59. Owen JG, Reddy BVB, Ternei MA, Charlop-Powers Z, Calle PY, Kim JH, et al. Mapping gene clusters within arrayed metagenomic libraries to expand the structural diversity of biomedically relevant natural products. Proc Natl Acad Sci U S A. 2013;110: 11797–11802.

60. Nazari B, Forneris CC, Gibson MI, Moon K, Schramma KR, Seyedsayamdost MR. sp. ATCC 55076 harbours the largest actinomycete chromosome to date and the kistamicin biosynthetic gene cluster. Medchemcomm. 2017;8: 780–788.

61. Kwun MJ, Cheng J, Yang SH, Lee D-R, Suh J-W, Hong H-J. Draft Genome Sequence of Ristocetin-Producing Strain Amycolatopsis sp. Strain MJM2582 Isolated in South Korea. Genome Announc. 2014;2. doi:10.1128/genomeA.01091-14

62. Banik JJ, Brady SF. Cloning and characterization of new glycopeptide gene clusters found in an environmental DNA megalibrary. Proc Natl Acad Sci U S A. 2008;105: 17273–17277.

63. Terlouw BR, Blin K, Navarro-Muñoz JC, Avalon NE, Chevrette MG, Egbert S, et al. MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. Nucleic Acids Res. 2023;51: D603–D610.

64. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al.

NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 2016;44: 6614–6624.

65. Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, et al. RefSeq: An update on prokaryotic genome annotation and curation. Nucleic Acids Res. 2018;46: D851–D860.

66. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res. 2016;44: D67–72.

67. Chen I-MA, Chu K, Palaniappan K, Pillay M, Ratner A, Huang J, et al. IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. Nucleic Acids Res. 2019;47: D666–D677.

68. Richardson L, Allen B, Baldi G, Beracochea M, Bileschi ML, Burdett T, et al. MGnify: the microbiome sequence data analysis resource in 2023. Nucleic Acids Res. 2023;51: D753–D759.

69. Sharrar AM, Crits-Christoph A, Méheust R, Diamond S, Starr EP, Banfield JF. Bacterial Secondary Metabolite Biosynthetic Potential in Soil Varies with Phylum, Depth, and Vegetation Type. MBio. 2020;11. doi:10.1128/mBio.00416-20

70. Lavy A, McGrath DG, Matheus Carnevali PB, Wan J, Dong W, Tokunaga TK, et al. Microbial communities across a hillslope-riparian transect shaped by proximity to the stream, groundwater table, and weathered bedrock. Ecol Evol. 2019;9: 6869–6900.

71. Nayfach S, Roux S, Seshadri R, Udwary D, Varghese N, Schulz F, et al. Author Correction: A genomic catalog of Earth's microbiomes. Nat Biotechnol. 2021;39: 521.

72. Haslinger K, Peschke M, Brieke C, Maximowitsch E, Cryle MJ. X-domain of peptide synthetases recruits oxygenases crucial for glycopeptide biosynthesis. Nature. 2015;521: 105–109.

73. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, et al. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res. 2021;49: W29–W35.

74. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. Microb Genom. 2021;7. doi:10.1099/mgen.0.000685

75. Kieser T. Practical Streptomyces Genetics. 2000.

76. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13: e1005595.

77. Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, Kautsar SA, Tryon JH, Parkinson EI, et al. A computational framework to explore large-scale biosynthetic diversity. Nat Chem Biol. 2020;16: 60–68.

78. Gilchrist CLM, Chooi Y-H. clinker & clustermap.js: automatic generation of gene cluster comparison figures. Bioinformatics. 2021;37: 2473–2475.

79. Heald SL, Mueller L, Jeffs PW. Actinoidins A and A2: structure determination using 2D NMR methods. J Antibiot (Tokyo). 1987;40: 630–645.

80. Vértesy L, Fehlhaber HW, Kogler H, Limbert M. New 4-oxovancosamine-containing

glycopeptide antibiotics from Amycolatopsis sp. Y-86,21022. J Antibiot (Tokyo). 1996;49: 115–118.

81. Sheldrick GM, Jones PG, Kennard O, Williams DH, Smith GA. Structure of vancomycin and its complex with acetyl-D-alanyl-D-alanine. Nature. 1978;271: 223–225.

82. Christensen SB, Allaudeen HS, Burke MR, Carr SA, Chung SK, DePhillips P, et al. Parvodicin, a novel glycopeptide from a new species, Actinomadura parvosata: discovery, taxonomy, activity and structure elucidation. J Antibiot (Tokyo). 1987;40: 970–990.

83. Sztaricskai F, Harris CM, Neszmelyi A, Harris TM. Structural studies of ristocetin A (ristomycin A): carbohydrate-aglycone linkages. J Am Chem Soc. 1980;102: 7093–7099.

84. Zmijewski MJ Jr, Briggs B, Logan R, Boeck LD. Biosynthetic studies on antibiotic A47934. Antimicrob Agents Chemother. 1987;31: 1497–1501.

85. Parenti F. Structure and mechanism of action of teicoplanin. J Hosp Infect. 1986;7 Suppl A: 79–83.

86. Singh SB, Jayasuriya H, Salituro GM, Zink DL, Shafiee A, Heimbuch B, et al. The complestatins as HIV-1 integrase inhibitors. Efficient isolation, structure elucidation, and inhibitory activities of isocomplestatin, chloropeptin I, new complestatins, A and B, and acid-hydrolysis products of chloropeptin I. J Nat Prod. 2001;64: 874–882.

87. McGahren WJ, Martin JH, Morton GO, Hargreaves RT, Leese RA, Lovell FM, et al. Structure of avoparcin components. J Am Chem Soc. 1980;102: 1671–1684.

88. Sanchez ML, Wenzel RP, Jones RN. In vitro activity of decaplanin (M86-1410), a new glycopeptide antibiotic. Antimicrob Agents Chemother. 1992;36: 873–875.

89. Bajusz D, Rácz A, Héberger K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? J Cheminform. 2015;7: 20.

90. Morgan HL. The generation of a unique machine description for chemical structures-A technique developed at chemical abstracts service. J Chem Doc. 1965;5: 107–113.

91. Durant JL, Leland BA, Henry DR, Nourse JG. Reoptimization of MDL keys for use in drug discovery. J Chem Inf Comput Sci. 2002;42: 1273–1280.

92. Röttig M, Medema MH, Blin K, Weber T, Rausch C, Kohlbacher O. NRPSpredictor2--a web server for predicting NRPS adenylation domain specificity. Nucleic Acids Res. 2011;39: W362–7.

93. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, et al. AntiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. 2019;47: W81–W87.

94. Rausch C, Weber T, Kohlbacher O, Wohlleben W, Huson DH. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). Nucleic Acids Res. 2005;33: 5799–5808.

95. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25: 3389–3402.

96. Megrian D, Taib N, Jaffe AL, Banfield JF, Gribaldo S. Ancient origin and constrained evolution of the division and cell wall gene cluster in Bacteria. Nat Microbiol. 2022;7: 2114–2127.

97. Stachelhaus T, Mootz HD, Marahiel MA. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem Biol. 1999;6: 493–505.

98. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30: 772–780.

99. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37: 1530–1534.

100. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14: 587–589.

101. Huson DH, Dezulian T, Klopper T, Steel MA. Phylogenetic super-networks from partial trees. IEEE/ACM Trans Comput Biol Bioinform. 2004;1: 151–158.

102. Huson DH. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics. 1998;14: 68–73.

103. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006;23: 254–267.

104. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol. 2017;2: 1533–1542.

105. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25: 1972–1973.

106. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009;25: 1422–1423.

107. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 2009;26: 1641–1650.

108. Tria FDK, Landan G, Dagan T. Phylogenetic rooting using minimal ancestor deviation. Nat Ecol Evol. 2017;1: 193.

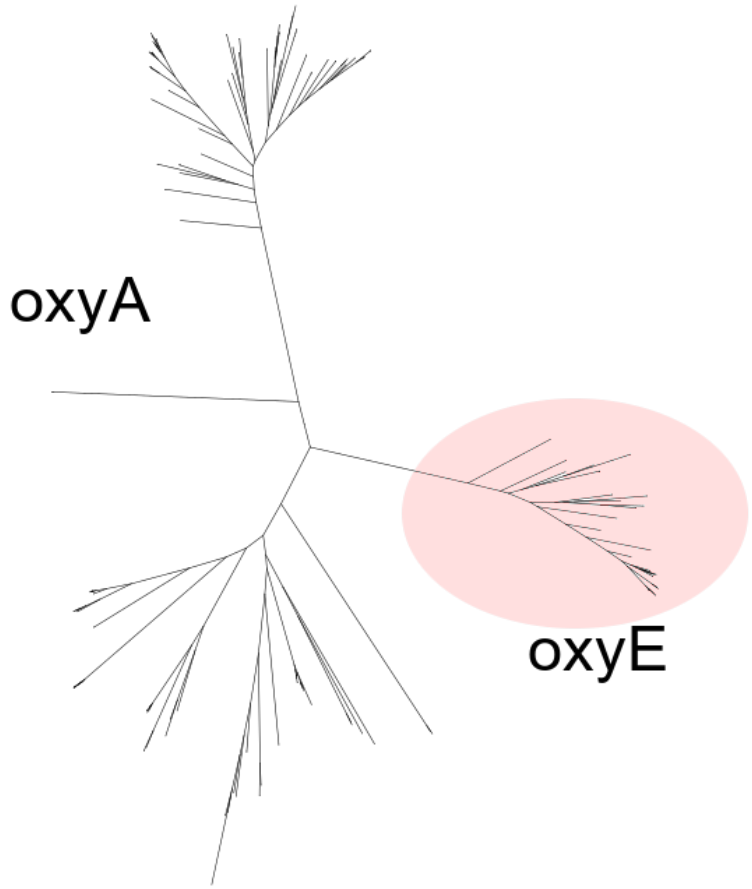
109. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49: W293–W296.

110. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 2024;52: W78–W82.

Supplementary Material

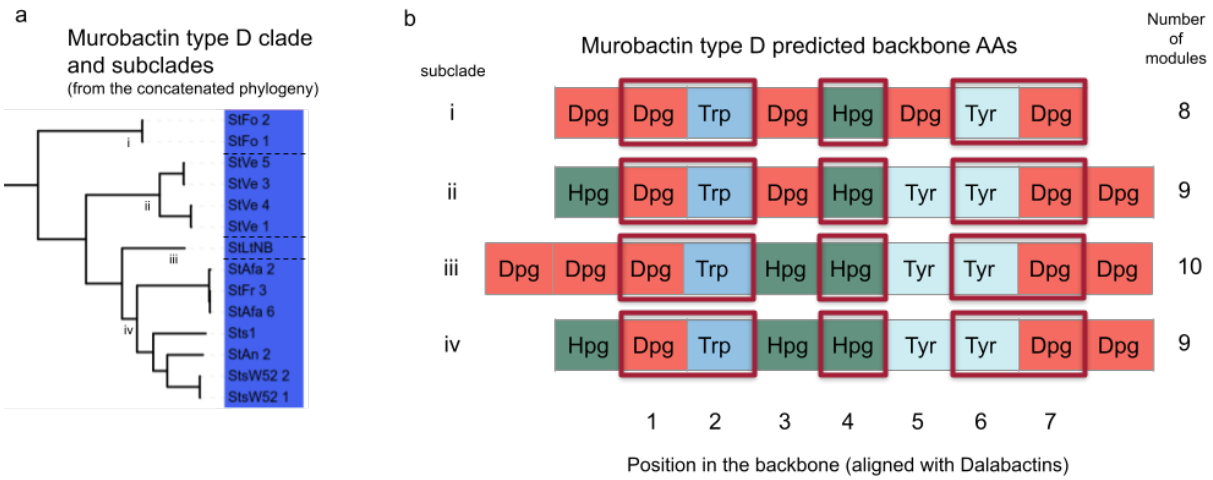
Supplementary Table 1: Most common gene categories found in dalabactin-synthesis (legacy types I-IV) encoding BGCs. The genes and their known functions are summarised in this table, organised into categories and subcategories based on their role. An example gene or domain is provided in the corresponding columns, as well as references.
Supplementary Table 2: SMILES of selected known xyclopeptide structures. Also, pairwise tanimoto similarity values.
Supplementary Table 3: Dataset of xyclopeptide BGCs and related metadata.
Supplementary Table 4: Results of selectivity code analysis of xyclopeptide BGC A-domains.
Supplementary Table 5: Table of orthologous groups and gene presence/absence table from Supplementary Figure 2.
Supplementary Table 6: Table of ML models used for the phylogenetic trees of genes and domains.
Supplementary Table 7: Results of congruence analysis.
Supplementary Data 1: SplitsTree6 session file, containing all (154) calculated phylogenetic trees, as well as the super network of Figure 3.
Supplementary Data 2: Fasta file of the concatenated MSA and newick files (unrooted with bootstrap values and MAD rooted) containing the concatenated phylogeny of xyclopeptides.
Supplementary Data 3: A-domain selectivity code-based concatenated MSA.
Supplementary Data 4: Tables of accession numbers of NCBI RefSeq and Genbank entries included in the X-domain search and the result of the search.
Supplementary Data 5: Custom scripts used for the analysis.
Supplementary Data 6: Sequencing report of *S. varsoviensis*.
Supplementary Data 7: Fasta files of all gene and domain MSAs (and trimmed MSAs) and newick files of the phylogenetic trees they were used to build.

Supplementary Figures

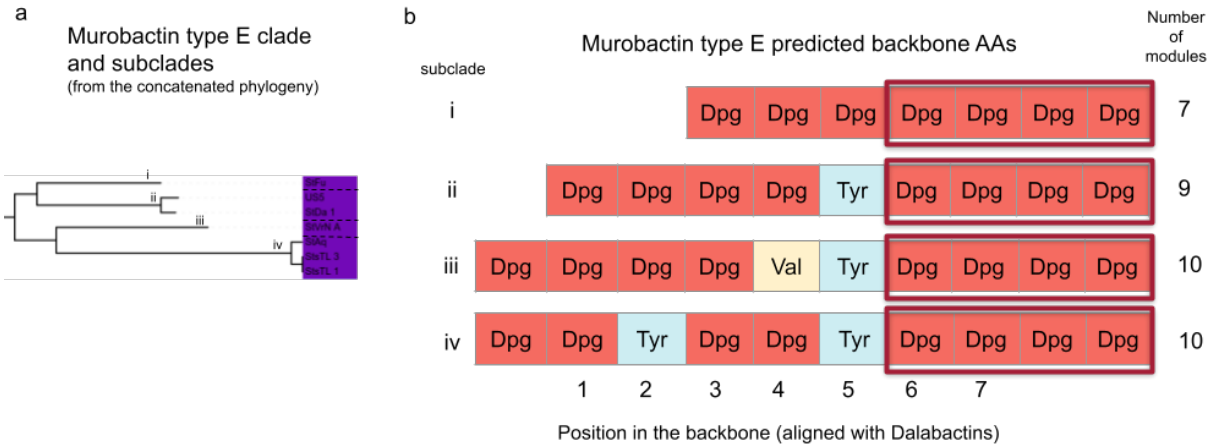


Supplementary Figure 1: Phylogenetic tree (unrooted) of the mixed zol orthogroup OG3, which includes P450 monooxygenases *oxyA* and *oxyE*. The clades of the two genes can be separated after annotation of the corresponding enzymes with known function. The known oxyE genes all belong to the clade in the coloured circle and were removed from OG3 and assigned to the artificial group OG3e. Trees of both OG3 and OG3e can be explored in **Supplementary Data 1**.

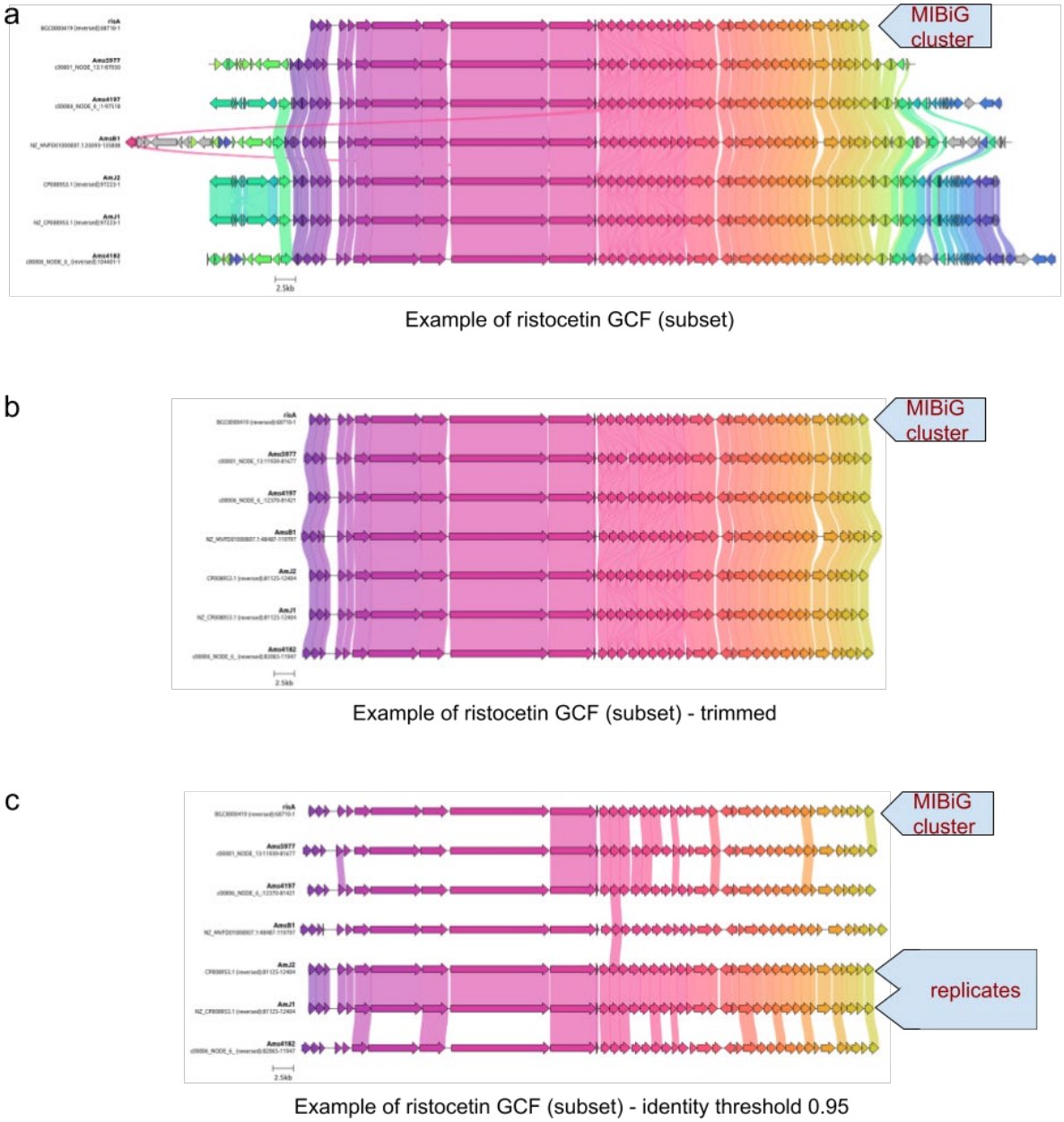
Supplementary Figure 2: High-resolution concatenated phylogeny of dalabactin and murobactin encoding BGCs (**Supplementary Data 2**, **Supplementary Table 3**). Information including types, backbone amino acids and gene content of the BGCs are plotted next to the tree.



Supplementary Figure 3: Murobactin type D subclades and backbone compositions. Panel a: clade of Murobactin type D, with marked subclades i-iv. BGC IDs are shown as leaf labels. Panel b: predicted backbone connected to its corresponding subclade. The red squares highlight the characteristic motif of the type.



Supplementary Figure 4: Murobactin type E subclades and backbone compositions. Panel a: clade of murobactin type E, with marked subclades i-iv. BGC IDs are shown as leaf labels. Panel b: predicted backbone connected to its corresponding subclade. The red squares highlight the characteristic motif of the type.



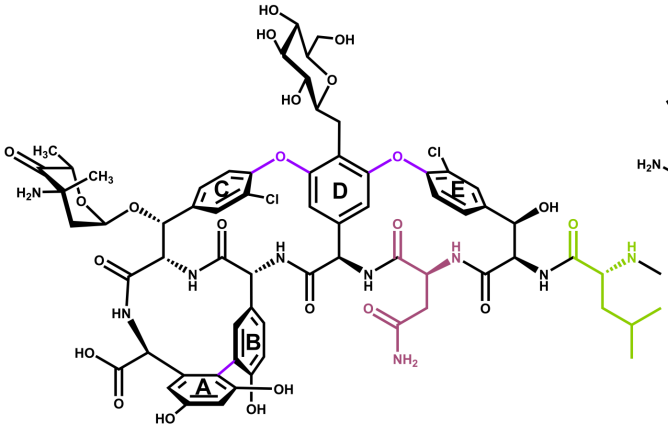
Supplementary Figure 5: Example of clinker-based trimming and dereplication. Panel a: subset of the ristocetin gene cluster family (GCF), visualised with clinker. Genes are coloured the same if they were placed in the same gene group based on similarity by clinker and are connected by coloured bands if their identity is higher than 0.5. The MIBIG cluster is the top one (marked) and the trimming will be based on it. Panel b: the same visualisation as panel a, after the trimming is finished. Now all BGCs belonging to this GCF are uniformly trimmed. Panel c:

Acknowledgements

AG and JR are grateful for the support of the Deutsche Forschungsgemeinschaft (DFG; Project ID No. 398967434-TRR 261). NZ and MA were supported by the German Center for Infection Research (DZIF) (TTU 09.716). NZ, MA, ES, NK, WW thank CMFI (Germany's Excellence Strategy–EXC 2124–390838134) for funding and structural support. NK, NZ, MC, OG and ES were supported through the JPIAMR project AVOGADRO, with funding for the German partners provided by the Bundesministerium für Forschung, Technologie und Raumfahrt (BMFTR, formerly BMBF; grant no. 01KI2503). This work was supported by Monash University and EMBL Australia. This research was conducted by the Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science (CE200100012) and funded by the Australian Government. We thank Ana Monica Daneliuc for her assistance in conducting the HMM search. We also thank L. do Presti for invaluable comments on the manuscript. HMM search of X-domain in RefSeq and Genbank proteomes was performed on the IBMI nodes.

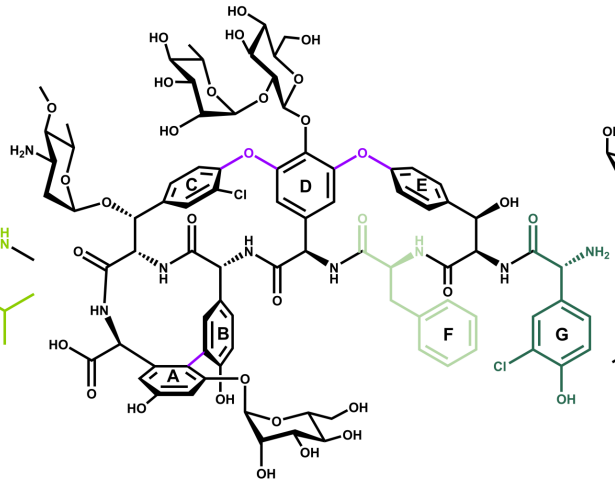
Conflict of interest

The following authors have no conflict of interest to declare: Athina Gavriilidou, Wolfgang Wohlleben, Nadine Ziemert, Noel Kubach, Martina Adamek,



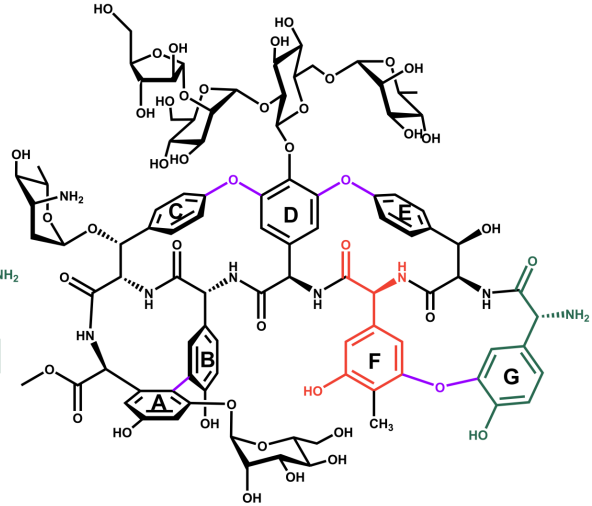
7-Dpg 6-Bht 5-Hpg 4-Hpg 3-Asn 2-Bht 1-Leu

Type I - balhimycin



7-Dpg 6-Bht 5-Hpg 4-Hpg 3-Phe 2-Bht 1-Hpg

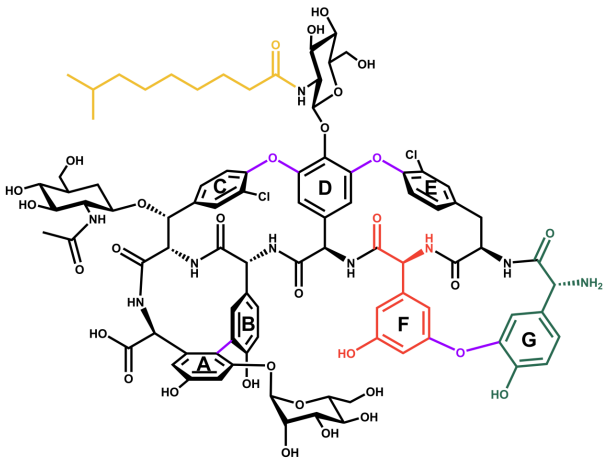
Type II - keratinimicin A



7-Dpg 6-Bht 5-Hpg 4-Hpg 3-Dpg 2-Bht 1-Hpg

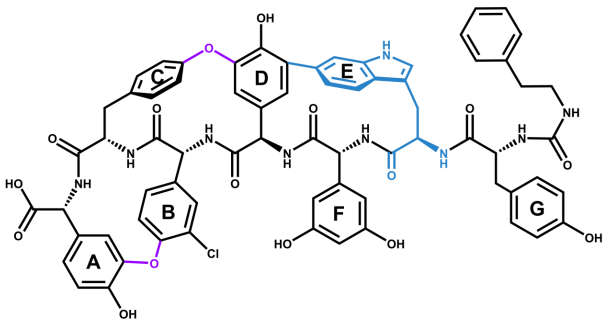
Type III - ristocetin A

Type	Side chains of AAs in position 1 and 3 of the backbone	Number of crosslinks	Additional unique characteristics
I	aliphatic (leu/ala & asn/glu)	3	-
II	aromatic (hpg & phe)	3	-
III	aromatic (hpg & dpg)	4	-
IV	aromatic (hpg & dpg)	4	Acylated sugar moiety
V	variable	variable	Trp in the backbone



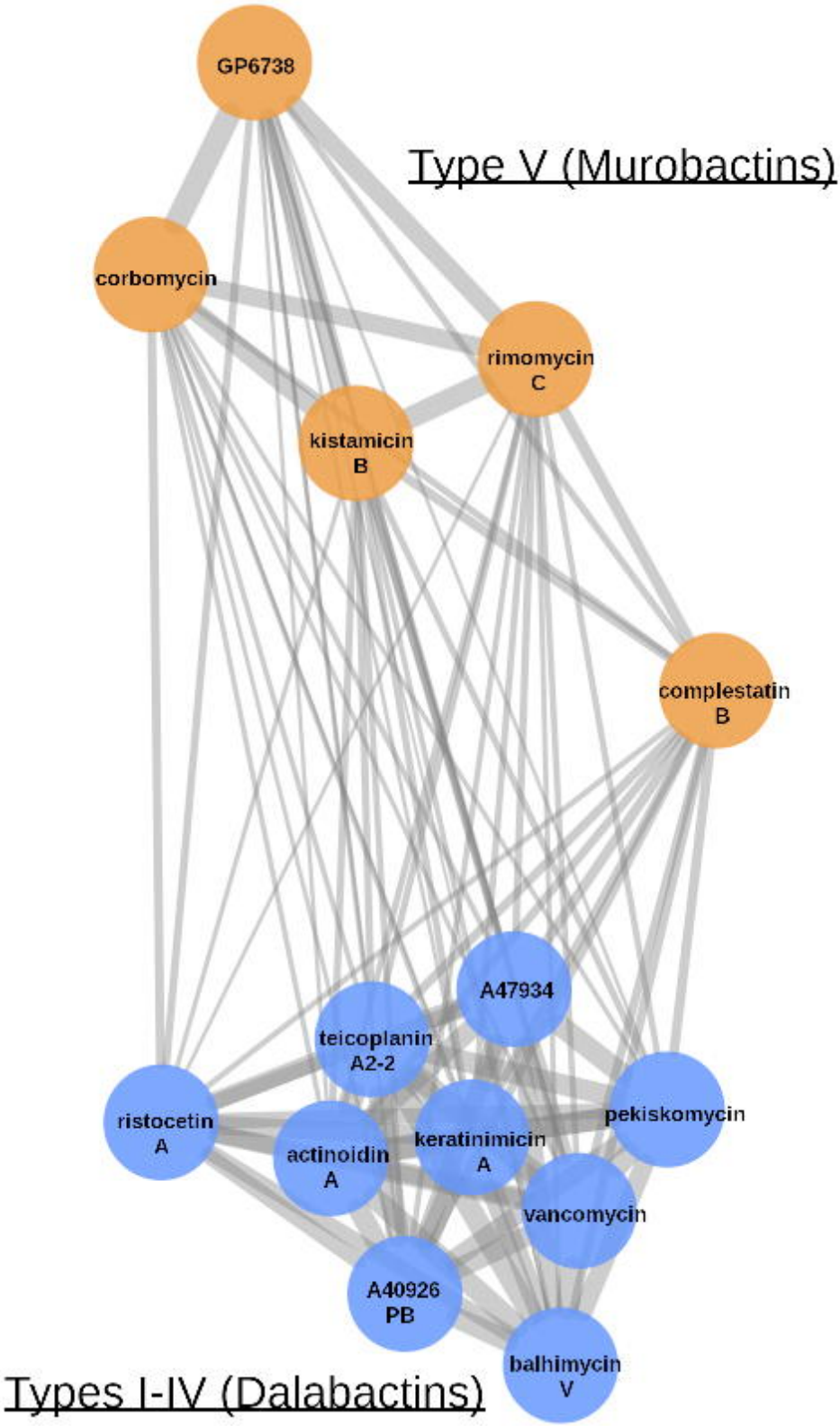
7-Dpg 6-Bht 5-Hpg 4-Hpg 3-Dpg 2-Tyr 1-Hpg

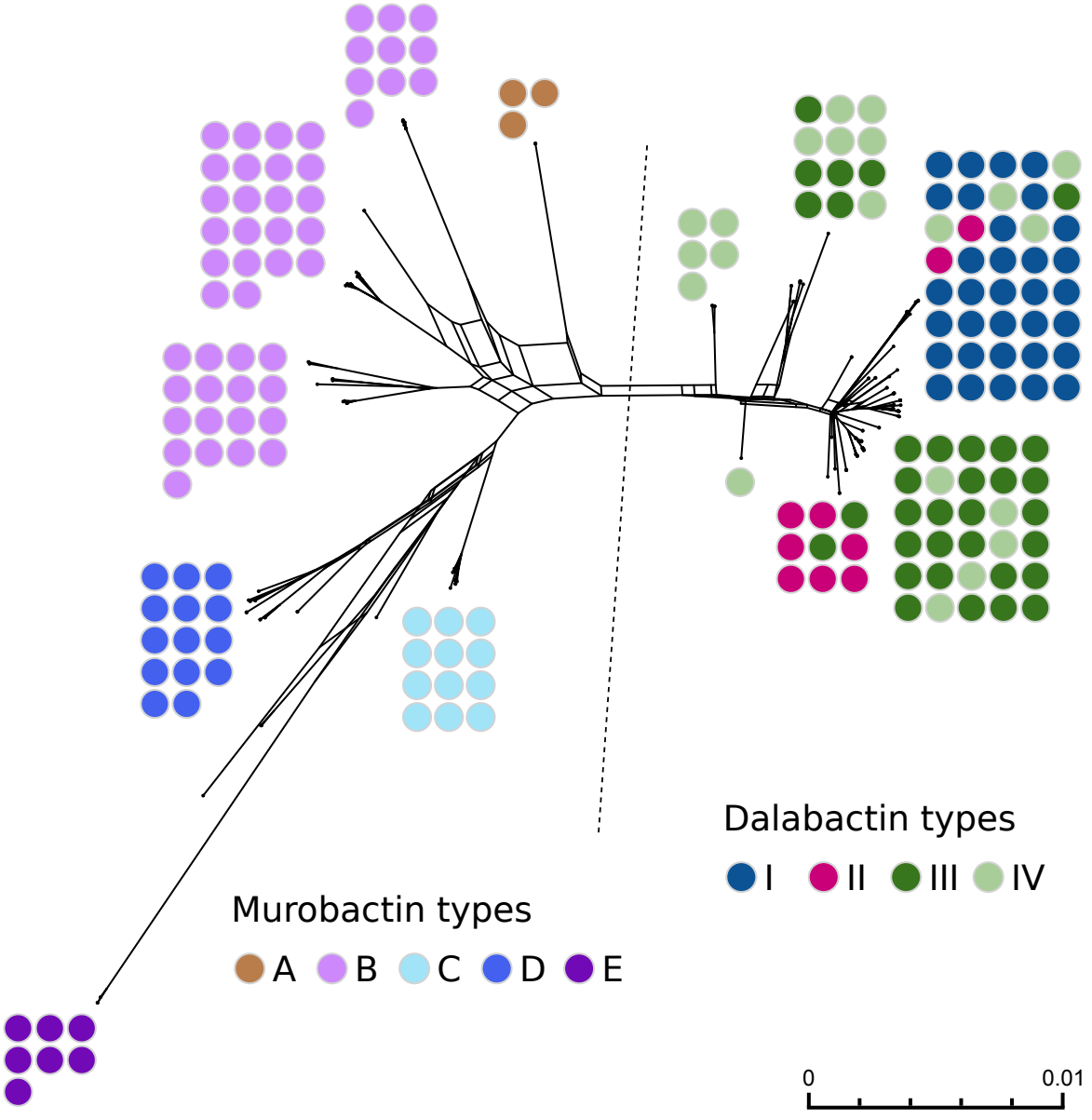
Type IV - teicoplanin A2-2

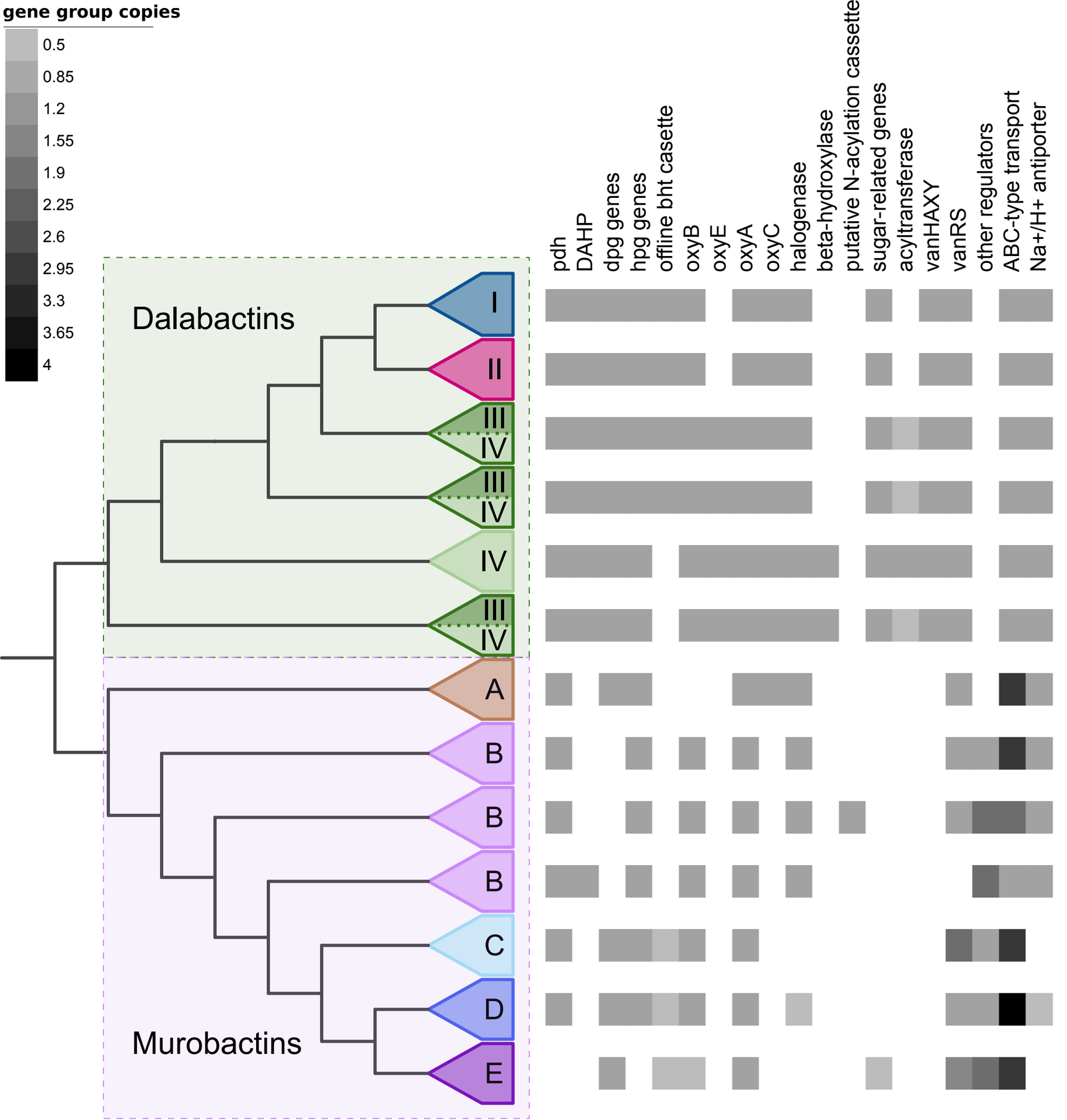


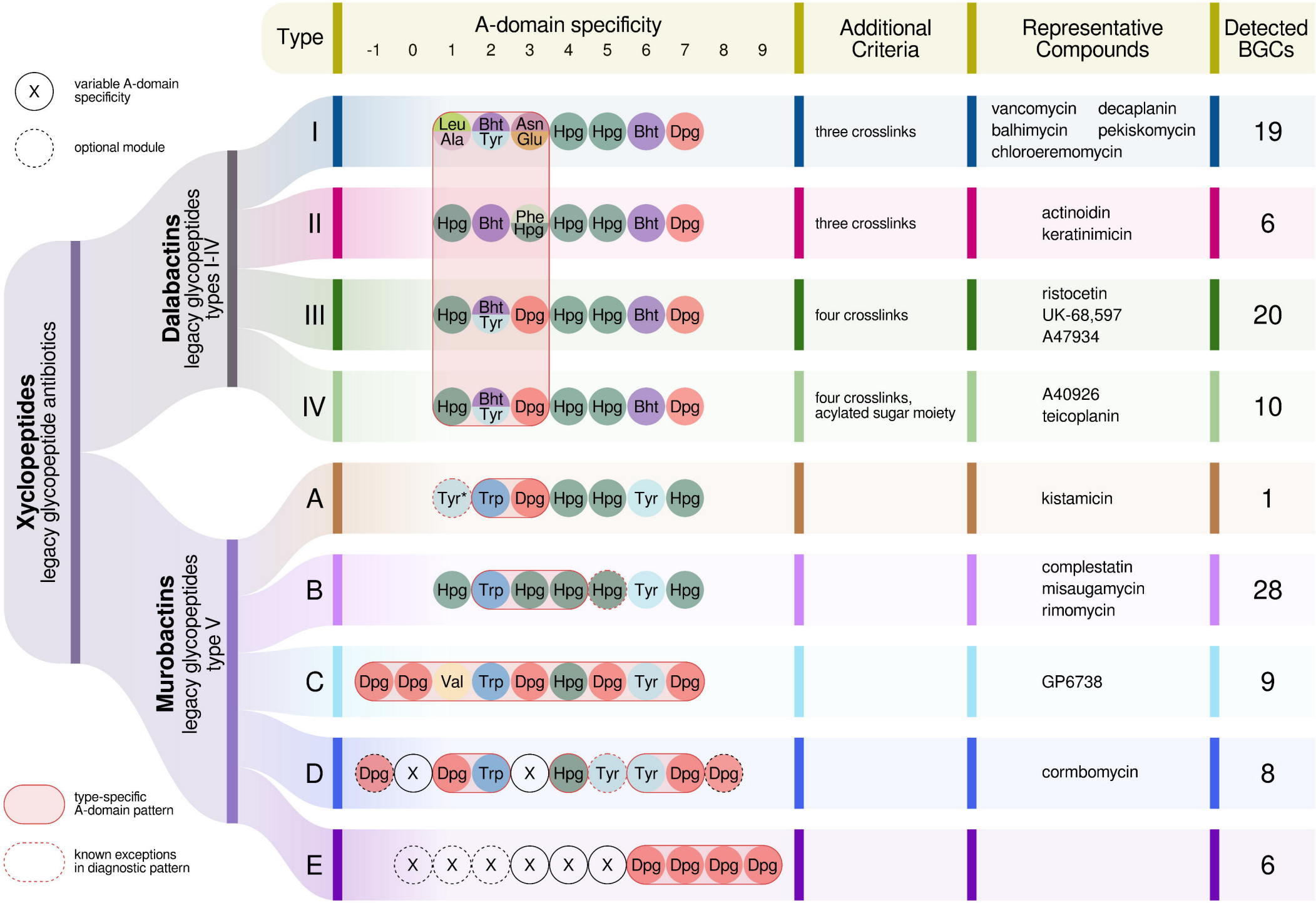
7-Hpg 6-Tyr 5-Hpg 4-Hpg 3-Dpg 2-Trp 1-Tyr

Type V - kistamicin B


GP6738
Type V (Murobactins)
corbomycin
rimomycin C
kistamicin B
complestatin B
A47934
teicoplanin A2-2
ristocetin A
actinoidin A
keratinimicin A
pekiskomycin
vancomycin
A40926 PB
balhimycin V
Types I-IV (Dalabactins)


Dalabactin types
I
II
III
IV
Murobactin types
A
B
C
D
E
0
0.01


gene group copies
0.5
0.85
1.2
1.55
1.9
2.25
2.6
2.95
3.3
3.65
4
Dalabactins
Murobactins
I
II
III
IV
III
IV
IV
III
IV
A
B
B
B
C
D
E
pdh
DAHP
dpg genes
hpg genes
offline bht casette
oxyB
oxyE
oxyA
oxyC
halogenase
beta-hydroxylase
putative N-acylation cassette
sugar-related genes
acyltransferase
vanHAXY
vanRS
other regulators
ABC-type transport
Na+/H+ antiporter


Type
A-domain specificity
-1 0 1 2 3 4 5 6 7 8 9
Additional Criteria
Representative Compounds
Detected BGCs
X variable A-domain specificity
optional module
Xyclopeptides
legacy glycopeptide antibiotics
Dalabactins
legacy glycopeptides types I-IV
Murobactins
legacy glycopeptides type V
I
Leu Ala
Bht Tyr
Asn Glu
Hpg Hpg Bht Dpg
three crosslinks
vancomycin decaplanin
balhimycin pekiskomycin
chloroeremomycin
19
II
Hpg Bht
Phe Hpg
Hpg Hpg Bht Dpg
three crosslinks
actinoidin
keratinimicin
6
III
Hpg
Bht Tyr
Dpg Hpg Hpg Bht Dpg
four crosslinks
ristocetin
UK-68,597
A47934
20
IV
Hpg
Bht Tyr
Dpg Hpg Hpg Bht Dpg
four crosslinks, acylated sugar moiety
A40926
teicoplanin
10
A
Tyr*
Trp Dpg Hpg Hpg Tyr Hpg
kistamicin
1
B
Hpg Trp Hpg Hpg Hpg Tyr Hpg
complestatin
misaugamycin
rimomycin
28
C
Dpg Dpg Val Trp Dpg Hpg Dpg Tyr Dpg
GP6738
9
D
Dpg X Dpg Trp X Hpg Tyr Tyr Dpg Dpg
cormbomycin
8
E
X X X X X X Dpg Dpg Dpg Dpg
6
type-specific A-domain pattern
known exceptions in diagnostic pattern