Automated genome mining predicts structural diversity and taxonomic distribution of peptide metallophores across bacteria

Zachary L. Reitz* [1,2], Bita Pourmohsenin* [3], Melanie Susman[4], Emil Thomsen[4], Daniel Roth[4], Alison Butler[4], Nadine Ziemert[# 3], Marnix H. Medema[# 1]

1. Bioinformatics Group, Wageningen University, 6708 PB Wageningen, the Netherlands
2. Department of Evolution, Ecology, and Marine Biology, University of California, Santa Barbara, CA 93117, USA
3. Interfaculty Institute of Microbiology and Infection Medicine, Institute for Bioinformatics and Medical Informatics, Tübingen, Germany
4. Department of Chemistry and Biochemistry, University of California, Santa Barbara, CA 93117, USA

* Contributed equally

Corresponding authors

Abstract

Microbial competition for trace metals shapes their communities and interactions with humans and plants. Many bacteria scavenge trace metals with metallophores, small molecules that chelate environmental metal ions. Metallophore production may be predicted by genome mining, where genomes are scanned for homologs of known biosynthetic gene clusters (BGCs). However, accurately detecting non-ribosomal peptide (NRP) metallophore biosynthesis requires expert manual inspection, stymieing large-scale investigations. Here, we introduce automated identification of NRP metallophore BGCs through a comprehensive algorithm, implemented in antiSMASH, that detects chelator biosynthesis genes with 97% precision and 78% recall against manual curation. We showcase the utility of the detection algorithm by experimentally characterizing metallophores from several taxa. High-throughput NRP metallophore BGC detection enabled metallophore detection across 69,929 genomes spanning the bacterial kingdom. We predict that 25% of all bacterial non-ribosomal peptide synthetases encode metallophore production and that significant chemical diversity remains undiscovered. A reconstructed evolutionary history of NRP metallophores supports that some chelating groups may predate the Great Oxygenation Event. The inclusion of NRP metallophore detection in antiSMASH will aid non-expert researchers and continue to facilitate large-scale investigations into metallophore biology.

Introduction

Across environments, microbes compete for a scarce pool of trace metals. Many microbes scavenge metal ions with small-molecule chelators called *metallophores*, which diffuse through the environment and chelate metal ions with high affinity.[1,2] A microbe possessing the right membrane transporters will be able to recognize and import a metallophore–metal complex, while other strains are unable to access the chelated metal ions. Thus, the metallophore secreted by one microbe can either support or inhibit growth of a neighboring strain, driving complex community dynamics in marine, freshwater, soil, and host environments.[3] The most well studied metallophores are the Fe(III)-binding *siderophores*, which have found applications in biocontrol, bioremediation, and medicine.[4] Two recent studies demonstrated that the disease suppression ability of a rhizosphere microbiome is strongly determined by whether or not the pathogen can use siderophores produced by the community; a microbiome can even encourage pathogen growth when a compatible siderophore is produced.[5,6] Compared to siderophores, other metallophore classes are relatively understudied, but they likely play equally important biological roles, as exemplified by recent reports of both commensal and pathogenic bacteria relying on zincophores to effectively colonize human hosts.[7,8]

Hundreds of unique metallophore structures have been characterized, each with specific chemical properties (e.g., effective pH range, hydrophobicity, and metal selectivity) and biological effects on other microbes (based on membrane transporter compatibility). Experimentally characterizing metallophores can be time-consuming and costly, and thus researchers often use genome mining to predict metallophore production *in silico*.[9] Taxonomy alone is not sufficient to predict what metallophores will be produced by a microbe, as production can vary significantly even within a single species.[10] Instead, metallophores must be predicted from each genome based on the presence of biosynthetic gene clusters (BGCs) that encode their biosynthesis. The majority of known metallophores are non-ribosomal peptides (NRPs), a broad class of natural products that also includes many antibiotics, antitumor compounds, and toxins. Specialized chelating moieties bind directly to the metal ion (in the case of siderophores, Fe^{3+}), while other amino acids in the peptide chain give the metallophore the required flexibility for chelation. Nearly all NRP metallophores contain one or more of the substructures shown in Fig. 1A: 2,3-dihydroxybenzoate (catechol, 2,3-DHB), hydroxamates, salicylate, β-hydroxyaspartate (β-OHAsp), β-hydroxyhistidine (β-OHHis), graminine, Dmaq (1,1-dimethyl-3-amino-1,2,3,4- tetrahydro-7,8-dihydroxy-quinoline), and the pyoverdine chromophore. Biosynthetic pathways are known for each of the chelating groups (Fig. 1B), and the presence of a chelator pathway may be used as a marker for metallophore production.

Mining genomes for metallophore BGCs has facilitated the discovery of chemically and biologically diverse metallophore systems; however, automated detection tools are still severely lacking.[9] The peptidic backbones of NRP metallophores are produced by non-ribosomal peptide synthetases (NRPSs), large multi-domain enzymes that activate and condense amino acids and other substrates in an assembly-line manner.[11] In the past two decades, a variety of bioinformatic tools have been developed to identify NRPS BGCs in a genome. One of the most popular is the secondary metabolite prediction platform antiSMASH, which uses a library of profile hidden Markov models (pHMMs) to identify (combinations of) enzyme-coding genes that are indicative of certain classes of specialized metabolite biosynthetic pathways.[12,13] For example, antiSMASH identifies an NRPS BGC region by the minimum requirement of a gene containing at least one condensation and one adenylation domain. NRP metallophore BGCs are technically detected by this rule as well; however, NRPSs also produce many other families of compounds, and additional manual annotation has still been required to identify NRP metallophore BGCs specifically. Accordingly, accurate prediction of BGCs encoding siderophores and other metallophores was limited to experts in natural product biosynthesis, and even experts cannot manually curate the thousands of BGCs produced by high-throughput metagenomic or comparative genomic analyses. To date, no global analysis of NRP metallophores has been performed, and thus the prevalence, combinatorics, and taxonomic distribution of different chelating groups are unknown.

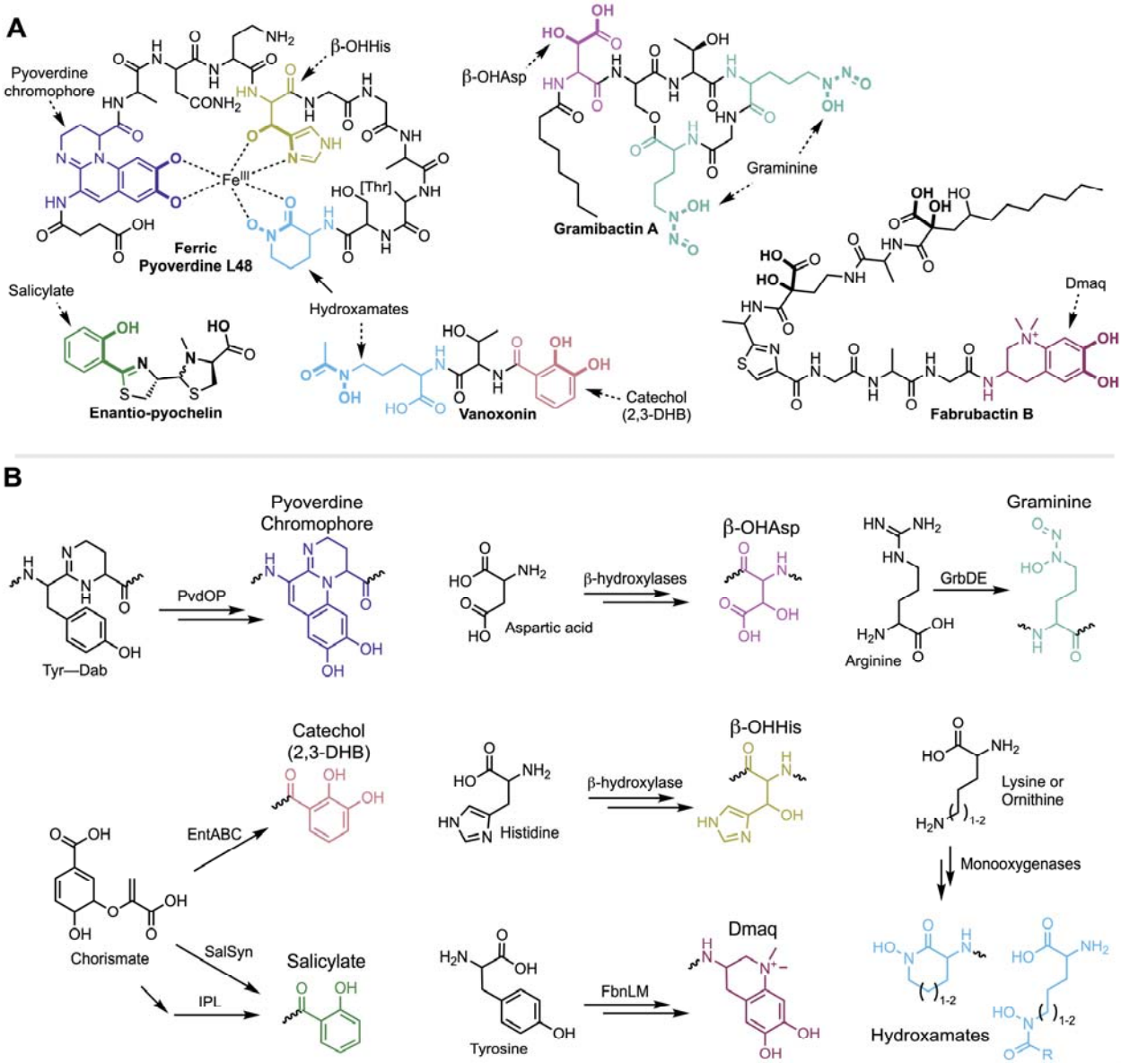


Figure 1. Chelating substructures found in bacterial NRP metallophores and their biosynthetic pathways. (A) Representative NRP metallophore structures. Nearly all known NRP metallophores contain one or more of the eight labeled chelating groups. Most chelating groups provide bidentate metal chelation, as shown for ferric pyoverdine L48. (B) Chelator biosynthesis pathways that form the basis for the new antiSMASH detection algorithm, as described in the text. The same chelator colors are used in each figure.

Here, we describe the development and application of a high-accuracy antiSMASH-integrated method for the automated detection of NRP metallophore BGCs, using the presence of chelator biosynthesis genes within NRPS BGCs as key markers for predicting metallophore production. The new detection rule was applied to 15,562 representative bacterial genomes, allowing us to take the first census of NRP metallophore production across bacteria. At least

25% of all NRPS clusters in these representative genomes code for the production of metallophores and significant biosynthetic diversity remains undiscovered. We then leveraged our computational analyses to guide characterization of siderophores from multiple bacterial taxa, finding structures that matched our genome-based predictions. By mapping NRP metallophore BGCs from 59,851 genomes to the Genome Taxonomic Database (GTDB) phylogeny, we identified myxobacterial and cyanobacterial metallophores as understudied and reconstructed a possible evolutionary history of the chelating groups.

Results

A chelator-based strategy for detection of NRP metallophore biosynthetic gene clusters

The specialized chelating moieties found in NRP metallophores are rarely found in other natural products, and thus we sought to automate metallophore BGC prediction by searching for genes encoding their biosynthesis. An extensive review of published NRP metallophore structures revealed that nearly all contain one or more of just eight chelator substructures (Fig. 1A). Protein domains responsible for their biosyntheses have been reported (Fig. 1B), and thus pHMMs could be constructed to allow detection of putative chelator biosynthesis genes. Generally, draft pHMMs were built from alignments of known and predicted NRP metallophore biosynthesis genes collected from literature, and cutoffs were manually determined (see Supplemental Discussion 1). The final multiple sequence alignments, pHMMs, and cutoffs are provided in the Supplemental Dataset.

A full description of each biosynthetic pathway detection strategy, including caveats and known limitations, is provided in Supplemental Discussion 1 and briefly summarized here. The profile HMMs implemented within antiSMASH are given in monospaced bold font. The biosynthetic cassette for 2,3-DHB is detected by an isochorismate synthase (**`EntC`**) and 2,3-dihydro-2,3- dihydroxybenzoate dehydrogenase (**`EntA`**).[14] Two salicylate biosynthesis pathways are detected by the presence of either an isochorismate pyruvate-lyase (**`IPL`**)[15] or a bifunctional salicylate synthase (**`SalSyn`**).[16] We also included detection of two condensation domain subtypes specific to catecholic and phenolic metallophores: VibH-like enzymes (**`VibH`**)[17,18] and tandem heterocyclization domains (**`Cy_tandem`**).[19] Peptidic hydroxamate pathways are detected by an ornithine (Orn) or Lys N-monooxygenase (**`Orn_monoox`** or **`Lys_monoox`**, respectively).[20] We could not accurately detect the vicibactin hydroxylase VbsO using a pHMM,[21] and so the characteristic acyl-hydroxyornithine epimerase **`VbsL`** is used to detect vicibactin biosynthesis.[21] We previously identified three families of siderophore-specific Fe(II)/α-ketoglutarate-dependent

enzymes responsible for β-OHAsp (**TBH_Asp** and **IBH_Asp**) or β-OHHis (**IBH_His**).[22] Based on the recent discovery of β-OHAsp-containing cyanochelins from cyanobacteria,[23] we have now identified two new clades that are putatively metallophore-specific and tentatively named **CyanoBH_Asp1** and **CyanoBH_Asp2**. The diazeniumdiolate-containing graminine may be detected by the presence of the cryptic necessary enzymes **GrbD** and **GrbE**.[24,25] The quinoline chelator Dmaq is detected by **FbnL** and **FbnM**, which initiate Dmaq biosynthesis.[26] The chromophore of pyoverdines is detected by the presence of a tyrosinase **PvdP** and/or an oxidoreductase **PvdO**.[27,28]

Several known chelating group pathways are not currently detected. Our detection strategy is limited to clades or combinations of biosynthetic enzymes that are distinct to NRP metallophore pathways. Several chelators are synthesized by the core NRPS and/or polyketide synthase (PKS) machinery and could not be detected without also retrieving many false positives, including NRPS-derived thiazol(id)ine and oxazol(id)ine heterocycles (see pyochelin, Fig. 1A) and PKS-derived 5-alkylsalicylate (e.g. in micocacidin[29]). We also did not include detection of a pathway currently only reported in fabrubactins that produces two α-hydroxycarboxylate chelating moieties (Fig. 1A**,** bolded atoms).[26] Finally, we have not yet designed detection rules for the recently discovered chelating groups 5-aminosalicylate of pseudonochelin[30] or 2-napthoate of ecteinamines;[31] however, we expect that their biosyntheses will be amenable to detection by the method used herein (Fig. S1). The NRP metallophore detection algorithm is publicly available in the antiSMASH web server and command line tool (https://antismash.secondarymetabolites.org, version 7 and upwards).

Validating antiSMASH NRP metallophore detection against manually curated BGCs

In order to assess the performance of our NRP metallophore BGC detection strategy, we manually predicted metallophore production among a large set of antiSMASH-predicted BGCs. A total of 758 NRPS BGC regions from 330 genera were annotated with default antiSMASH v6.1 and inspected for known markers of metallophore production, including genes encoding transporters, iron reductases, chelator biosynthesis, and known metallophore NRPS domain motifs. We thus manually classified 176 BGC regions (23%) as metallophore BGCs (Supplemental Table 2). The new antiSMASH detection rule was applied to the same BGC regions, resulting in 145 putative metallophore BGC regions (F1 = 0.86; Table 1 and Supplemental Table 2). Nine metallophore BGC regions were only detected by antiSMASH. Upon reinvestigation, four were determined to likely represent genuine metallophore BGC regions missed during manual analysis, leaving only five putative false positives in which

seemingly unrelated genes matched the pHMMs (97% precision). Conversely, a total of 40 metallophore BGC regions could only be detected manually (78% recall). In the majority of false negatives, NRP metallophore BGCs were missed because chelator biosynthesis genes, on which the detection strategy is based, were not present in the cluster. In 21 cases, genes encoding catechol, salicylate, or hydroxamate biosynthesis were located elsewhere in the genome. In ten cases, chelator biosynthesis pathways were not found anywhere in the genome; these clusters may be non-functional fragments, rely on exogenous precursors (as seen in equibactin biosynthesis[32]), or have evolved to use novel chelator biosyntheses. Two of the false negatives encoded the 5-alkylsalicylate PKS that is currently undetectable, as described above. Finally, seven manually assigned NRP metallophore BGC regions had no genes corresponding to known chelator pathways (Supplemental Table 2); if correctly annotated, they may represent novel structural classes. In one particularly promising case, a salicylate pathway appears to have been replaced with a partial menaquinone pathway to produce a putative 1,4-dihydroxy-2-naphthoate chelating group (Supplemental discussion 2).

	Performance metrics[a]			Number of NRP metallophore BGC regions detected in representative bacterial genomes[b]		
	Precision	Recall	F1[c]	Complete NRPS regions *n*=11,704	Partial NRPS regions *n*=8,403	Total NRPS regions *n*=20,107
AntiSMASH rule	0.97	0.78	0.86	2,485 (21%)[d]	725 (8.6%)	3,210 (16%)
Transporter genes	0.93	0.56	0.69	1,723 (15%)	376 (4.5%)	2,099 (10%)
Either/or ensemble	0.92	0.88	0.90	2,948 (25%)	855 (10%)	3,803 (19%)

Table 1. Summary of NRP metallophore BGC detection, comparing the chelator-based rule newly implemented in antiSMASH, the transporter-based method of Crits-Christoph *et al.*,[41] and a combined either/or ensemble. [a] Detection methods were each tested on a set of 758 manually annotated NRPS BGC regions (180 true positives). Full results are given in Supplemental Table 2. [b] Detection methods were applied to 15,562 NCBI RefSeq representative bacterial genomes. The full results are given in Supplemental Table 3. A region is "complete" if it is not on a contig edge, as determined by antiSMASH. An additional 54 BGC regions were detected as NRP metallophores without meeting the requirements for the antiSMASH NRPS rule. [c] F1 score is equal to 2×(Precision×Recall)/(Precision+Recall). [d] Percentages indicate the fraction of NRPS regions that were predicted to encode NRP metallophores.

AntiSMASH outperforms transporter-based detection, although both techniques are complementary

Crits-Christoph et al. found that the presence of transporters could be used to predict siderophore BGCs among other NRPS clusters.[33] Specifically, the Pfam families for TonB-

dependent receptors, FecCD domains, and periplasmic binding proteins (PF00593, PF01032, and PF01497, respectively) were determined to be highly siderophore-specific, and the authors used the presence of two of the three domains to predict a “siderophore-like” BGC region (metallophores that transport other metals were also coded as siderophores in their dataset). We used a modified version of antiSMASH to detect the three transporter families among the 758 manually annotated NRPS BGC regions (Table 1 and Supplemental Table 2). In total, the transporter-based method detected 108 metallophore clusters (F1 = 0.69), including eight putative false positives (93% precision), and had 80 false negatives (56% recall). One false positive was noted in the manual annotation as a likely “cheater”: while several *Bordetella* genomes encode the synthesis of a putative graminine-containing metallophore, *B. petrii* DSM 12804 has retained only the transporter genes alongside a small fragment of the NRPS. In the seven other false positives, BGC regions appeared to coincidentally contain transporter genes in their periphery, as they were not conserved in homologous NRPS clusters. In one case, the triggering genes were part of a putative vitamin B12 import and biosynthesis locus. Combining the two methods in an either/or ensemble approach slightly improved overall performance *versus* the antiSMASH rule alone, achieving 92% precision, 88% recall, and an F1 score of 0.90 (Table 1).

Charting NRP metallophore biosynthesis across bacteria

The implementation of NRP metallophore BGC detection into antiSMASH allowed us to take the first bacterial census of NRP metallophore biosynthesis. The finalized detection rule was applied to 15,562 representative bacterial genomes from NCBI RefSeq (25 June 2022). In total, 3,264 NRP metallophore BGC regions were detected (Table 1 and Supplemental Table 3), including 54 Type II (non- or semi-modular[34]) NRPS regions that would otherwise not be detected by antiSMASH, such as BGCs for acinetobactin and brucebactin.[35,36] NRP metallophores comprised 16% of all NRPS BGC regions in the genomes. Among complete regions (not located on a contig edge), 21% of NRPS BGC regions were classified as NRP metallophores, compared to just 8.6% of incomplete NRPS regions. This is consistent with previous reports that low-quality, fragmented genomes result in low-quality BGC annotations in antiSMASH.[37] The transporter-based approach predicted siderophore activity for 15% of complete NRPS regions, including 463 BGC regions without detectable chelator genes; when the two methods are combined, over 25% of NRPS BGCs are predicted to produce NRP metallophores (Table 1). Only complete NRP metallophore BGC regions detected by antiSMASH were used for downstream analyses.

Frequency and hybridization of NRP metallophore chelating groups

Complete NRP metallophore regions from the representative genomes were categorized by the type(s) of chelator biosynthesis genes detected within (Fig. 2). Hydroxamates and catechols were the most common pathways, present in 44% and 36% of BGC regions, respectively. In contrast, β-OHHis, graminine, and Dmaq biosyntheses were rare in representative genomes, each present in less than 2% of detected regions. About 20% of regions contained genes for at least two pathways and putatively produce a hybrid metallophore. Only 42 BGC regions (1.7%) contained three different chelating groups: each encoded genes for the pyoverdine chromophore, a hydroxamate, and either β-OHAsp or β-OHHis. The proportion of hybrid metallophores is likely higher than estimated here. As described above, some chelating moieties could not be captured by the pHMM-based rule. Furthermore, metallophore biosynthesis may require genes from multiple BGCs. Pyoverdine genes may be located in up to five different loci,[38] and all 56 regions with only the pyoverdine chromophore pathway are expected to produce hybrid siderophores. Representative characterized siderophores that contain the chelator combinations in our dataset are shown in Fig. 3.

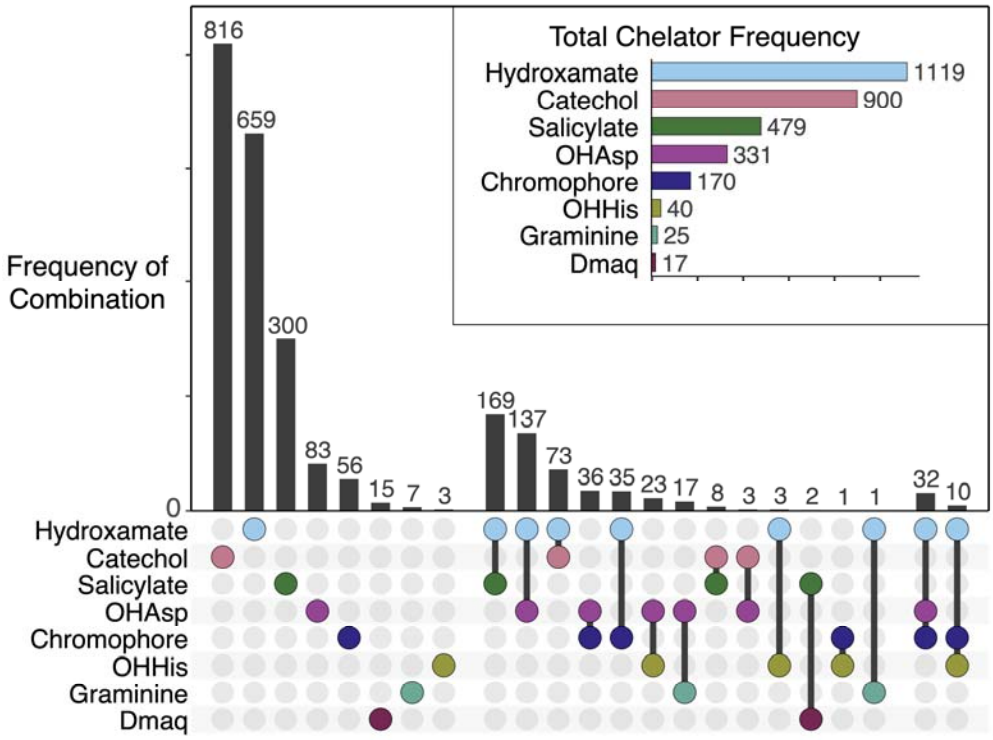
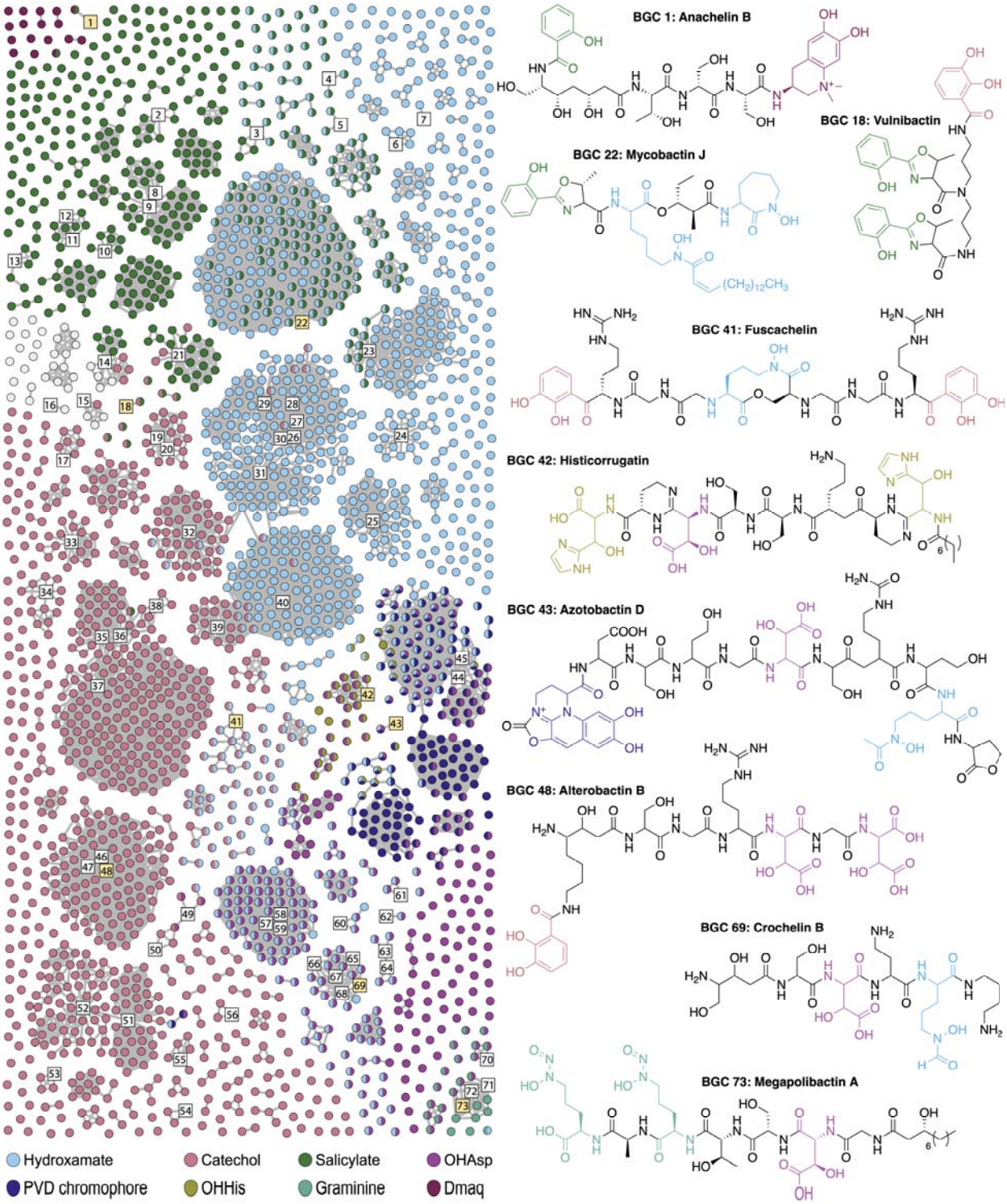


Figure 2. An upset plot of chelator frequency among 2,489 complete NRP metallophore BGC regions from RefSeq representative genomes. An additional 38 BGC regions were detected by metallophore-specific NRPS domains (VibH-like or tandem Cy) rather than chelator biosynthesis, and may produce catechol and/or salicylate metallophores using biosyntheses encoded elsewhere in the genome.

The most widespread NRP metallophore families have likely been found, yet significant diversity remains unexplored

Different species of bacteria can contain highly similar metallophore BGCs. To gauge the biosynthetic diversity of the putative NRP metallophores (and thereby the structural diversity), the complete BGC regions were organized into a sequence similarity network using BiG-SCAPE, which clusters BGCs based on their shared gene content and identity. An additional 75 reported NRP metallophore BGCs were included as reference nodes (Supplemental Table 1), and a distance cutoff of 0.5 was chosen to allow highly similar reference BGCs to form connected components (gene cluster families; GCFs) in the network. The final network, colored and organized by chelator type, is presented in Fig. 3. The majority of BGC regions (57%) clustered with the reference BGCs in just 45 GCFs, suggesting that many of the most widespread NRP metallophore families with known chelating groups already have characterized representatives (Fig. S4). However, 1093 BGC regions did not cluster with any reference BGC, forming 619 separate GCFs in the network (93% of all GCFs). Some of these may encode orphan metallophores previously isolated from unsequenced strains, or be similar to known BGCs that were not included in our non-exhaustive literature search. Nevertheless, significant NRP metallophore structural diversity remains undiscovered, particularly among the 484 BGC regions distinct enough to form isolated nodes in the network.

Figure 3 (Next page). BiG-SCAPE similarity network of complete NRP metallophore BGC regions from RefSeq representative genomes. Numbered square nodes indicate published BGCs, as given in Supplemental Table 1. Select hybrid metallophore BGC nodes are highlighted yellow, and their corresponding structures are shown. Nodes are colored by the type(s) of chelator biosynthesis detected therein. BGC regions colored light gray contain only metallophore-specific NRPS domains (VibH-like or tandem Cy) and may produce catechol and/or salicylate metallophores using biosyntheses encoded elsewhere in the genome. The network was constructed in BiG-SCAPE v1.1.2 using 2,596 BGC regions as input, including 78 reference BGCs, and a distance cutoff of 0.5.


BGC 1: Anachelin B
BGC 18: Vulnibactin
BGC 22: Mycobactin J
BGC 41: Fuscachelin
BGC 42: Histicorrugatin
BGC 43: Azotobactin D
BGC 48: Alterobactin B
BGC 69: Crochelin B
BGC 73: Megapolibactin A
Hydroxamate
Catechol
Salicylate
OHAsp
PVD chromophore
OHHis
Graminine
Dmaq

Chemical identification of genome-predicted siderophores

To showcase how our large-scale automated genome mining methodology can be used to effectively predict functional metallophore biosynthetic pathways, we experimentally characterized the metallophores of three bacterial strains. Two strains belong to genera that have no reported metallophores: *Buttiauxella brennerae* DSM 9396 was predicted to produce enterobactin (Fig. 4a), and *Terasakiispira papahanaumokuakeensis* DSM 29361 was predicted to produce both marinobactin(s) (Fig. 4a) and enantio-pyochelin (Fig. 1a). The third strain, *Pseudomonas brassicacearum* DSM 13227, was selected because its genome contains a BGC that clustered with the histicorrugatin reference BGC in the BiG-SCAPE network (Fig. 3, node 42). We predicted that the BGC may encode the biosynthesis of ornicorrugatin (Fig. 4a),[39] a previously reported siderophore with no known BGC. A fragmented pyoverdine BGC was also present in the strain's genome, which was predicted to produce the known siderophore pyoverdine A214 (Fig. 4a).[40,41]

Each strain was grown in low-iron conditions to induce siderophore production, then organic compounds were extracted from the culture supernatants using adsorbant resin prior to spectral analyses by electrospray ionization mass spectrometry (ESI-MS) and ESI-MS/MS; full details are provided in the Supplemental Methods and Results. From *B. brennerae,* we identified four catecholic compounds (Fig. 4B): the predicted enterobactin (Fig. 4A), as well as the enterobactin fragments 2,3-DHB–Ser (DHBS), $(DHBS)_2$ and linear $(DHBS)_3$. The crude extract of *T. papahanaumokuakeensis* indeed contained molecular ions consistent with marinobactins A-E (Fig. 4A and C). Tandem ESI-MS/MS yielded expected fragmentation patterns for marinobactins A-D, while the peak at *m/z* 988.5421, putatively marinobactin E, was low abundance and did not give a clear spectrum. No peaks consistent with enantio-pyochelin (*m/z* 324.4; Fig. 1a) could be observed. From *P. brassicacearum*, we identified both siderophores predicted from the BGC analyses: ornicorrugatin and pyoverdine A214 (Fig. 4A and D). Fragmentation patterns closely matched those previously reported.[39,40]

Thus, our method was able to successfully identify the putative BGC for the orphan siderophore ornicorrugatin and also correctly predict the potential to produce known siderophores by taxa that were not yet studied for their metallophore biosynthetic capacities.

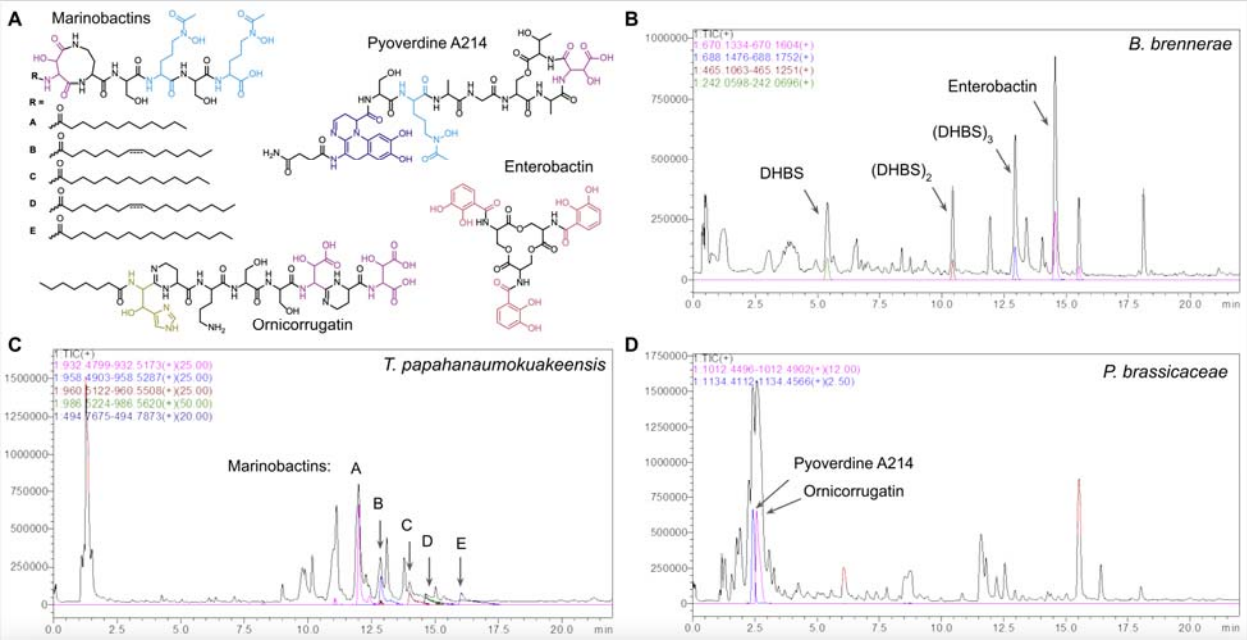


Figure 4. Identification of siderophores predicted from genome mining. (A) Chemical structures of marinobactins A-E,[42] produced by *Terasakiispira papahanaumokuakeensis* DSM 29361; enterobactin,[43] produced by *Buttiauxella brennerae* DSM 9396; and pyoverdine A214[40] and ornicorrugatin,[39] both produced by *Pseudomonas brassicacearum* DSM 13227. The position and orientation of the fatty acid desaturation in marinobactins B and D was not determined in this work. (B-D) High pressure liquid chromatography / high-resolution mass spectrometry (HPLC-HRMS) total ion chromatograms of culture supernatant extracts, overlaid with extracted ion chromatograms for siderophore features. Additional details and spectra are provided in the Supplemental Methods and Results.

Taxonomic distribution of NRP-metallophores

We investigated the taxonomic distribution and evolution of NRP siderophore biosynthesis within the bacterial kingdom by applying our antiSMASH detection rule to 59,851 representative bacterial genomes from the Genome Taxonomy Database (GTDB).[44] Among these, 4,098 genomes (6.8%) were predicted to contain at least one NRP metallophore BGC. A total of 5,366 BGC regions were detected, representing 14% of all detected NRPS regions. Taxonomic distribution analysis using the GTDB phylogeny highlighted the uneven prevalence of NRP-metallophores across bacterial phyla (Table 2). Proteobacteria and Actinomycetota were overrepresented in the GTDB representatives, together accounting for 89% of all detected NRP metallophore regions. After correcting for the number of representative genomes in each

phylum, NRP metallophore BGCs were most abundant in Actinomycetota, with 23% of genomes containing at least one detectable region. Proteobacteria, Cyanobacteria, and Myxococcota each had similar proportions of genomes with NRP metallophore BGCs; however, due to biased coverage in the GTDB database, 49% of the detected BGC regions were from Proteobacteria, compared to only 4% and 1.1% found in Cyanobacteria and Myxococcota. Thus, we expect that further sequencing efforts directed at these two phyla will yield many new NRP metallophore BGCs.

Phylum	**Number of detected NRP metallophore BGC regions**	**Percentage of total detected NRP-met regions**	**Proportion of genomes with ≥1 NRP-met regions**
Proteobacteria	2439	49%	2042/16536 (12%)
Actinomycetota	1986	40%	1561/6931 (23%)
Cyanobacteria	200	4.0%	176/1318 (13%)
Firmicutes_I	192	3.9%	191/4013 (4.8%)
Myxococcota	55	1.1%	52/418 (12%)
Firmicutes	28	0.6%	28/9026 (0.3%)
Chloroflexota	18	0.4%	14/1317 (1.1%)
Nitrospirota	16	0.3%	15/307 (4.9%)
Acidobacteriota	9	0.2%	9/836 (1.1%)
Desulfobacterota	5	0.1%	5/847 (0.6%)
Verrucomicrobiota	2	<0.1%	2/1304 (0.2%)
Planctomycetota	1	<0.1%	1/1034 (0.1%)
Bdellovibrionota	1	<0.1%	1/248 (0.4%)
Gemmatimonadota	1	<0.1%	1/345 (0.3%)

Table 2. Taxonomic distribution of 4,953 NRP-metallophore BGC regions detected in 59,851 GTDB representative bacterial genomes. Phylum nomenclature is preserved from GTDB r207. An additional 413 BGC regions with “unknown” taxonomy are not included here. Phyla not listed had zero detected regions.

To map the distribution of NRP-metallophore producers across the bacterial kingdom, we employed Relative Evolutionary Divergence (RED) values, a framework proposed by Parks et al. and utilized within the GTDB.[45] Building on this, Gavriilidou et al. defined REDgroups—phylogenetically consistent clusters based on RED values—that provide a standardized framework analogous to genera.[46] Unlike traditional genera, which can vary significantly in their evolutionary distances, REDgroups offer greater consistency in evolutionary relationships among their members. This framework allowed us to summarize the data as the average number of NRP-metallophore BGC regions per genome within each group, enabling effective visualization and more equitable comparative analyses of biosynthetic potential across bacterial lineages. By collapsing the GTDB tree to the REDgroup level, we annotated each group with the average number of putative NRP-metallophore clusters (Fig. 5, Ring C). The analysis revealed that 16% of REDgroups encoded detected NRP-metallophores, and within each REDgroup, the number of NRP-metallophores was relatively consistent (standard deviation: 0.3425). This observation aligns with the findings of Gavriilidou et al., who demonstrated that BGC diversity is consistent at the genus level.[46] While most REDgroups with NRP-metallophores averaged one per genome, several REDgroups, particularly within Actinomycetota, Proteobacteria, and Cyanobacteria exhibited higher averages, with some exceeding two per genome (Fig. 5, Ring C). These results reveal lineage-specific patterns in siderophore biosynthesis and highlight the utility of REDgroups as an alternative to traditional taxonomic units.

Evolution of Gene Families and Phylogenetic Reconciliation to Uncover the Evolutionary History of NRP-Metallophores

To investigate the evolution and origins of NRP-metallophores, we conducted a detailed phylogenetic analysis of each chelator group. Elucidating the evolutionary history of bacterial gene families is complicated by gene duplications, horizontal gene transfers (HGTs), and deletions that cause discordance between the bacterial species phylogeny and each chelator gene phylogeny. To reconcile the trees, we used the software package eMPRess, which infers the most likely series of duplication, HGT, and deletion events (maximum parsimony reconciliation) to reconstruct the evolutionary history of the gene family.[47] We first extracted non-fragmented BGC regions from the GTDB representative genomes, then clustered them with BiG-SCAPE to yield 1,108 representative BGCs. From these BGCs, we extracted chelator [48] biosynthesis genes and reconstructed gene trees, which were then compared to the GTDB species tree with eMPRess.[47] We validated the eMPRess results by analyzing two gene families

with the likelihood-based tool AleRax,[48] which produced duplication, transfer, and loss (DTL) patterns consistent with those obtained from eMPRess.

Estimates for the origins and early HGTs of the chelating groups are presented in the center of Fig. 5. Reconciliation indicates that the most wide-spread chelating groups—catechols, hydroxamates, and salicylates—are among the most ancient. Genes for producing 2,3-DHB may have originated in a common ancestor of Actinobacteriota (ca. 2.7 Ga, according to rough estimates from TimeTree[49]) and were then transferred stepwise to Proteobacteria and to Firmicutes. Salicylate biosynthesis genes have an estimated origin in a common ancestor of Gammaproteobacteria (ca. 1.9 Ga [49]), with early HGT to Cyanobacteria and Actinobacteriota. Hydroxamate NRP metallophores appear to have originated in the common ancestor of Alpha- and Gammaproteobacteria (ca. 2.3 Ga [49]) and were transferred into Actinobacteria, while Lys-based hydroxamates evolved in Actinobacteriota. The other chelator groups display a more phylum-specific distribution, with HGT predominantly occurring within the same phylum (see Supplemental Dataset, empress_reconciliations). Dmaq is predicted to be among the oldest chelating groups and may have been produced by the common ancestor of Cyanobacteria (ca. 2.7 Ga [49]), while the pyoverdine chromophore, exclusively observed within the order Pseudomonadales, likely represents one of the most recent siderophore biosynthetic pathways (ca. 1.2 Ga [49]).

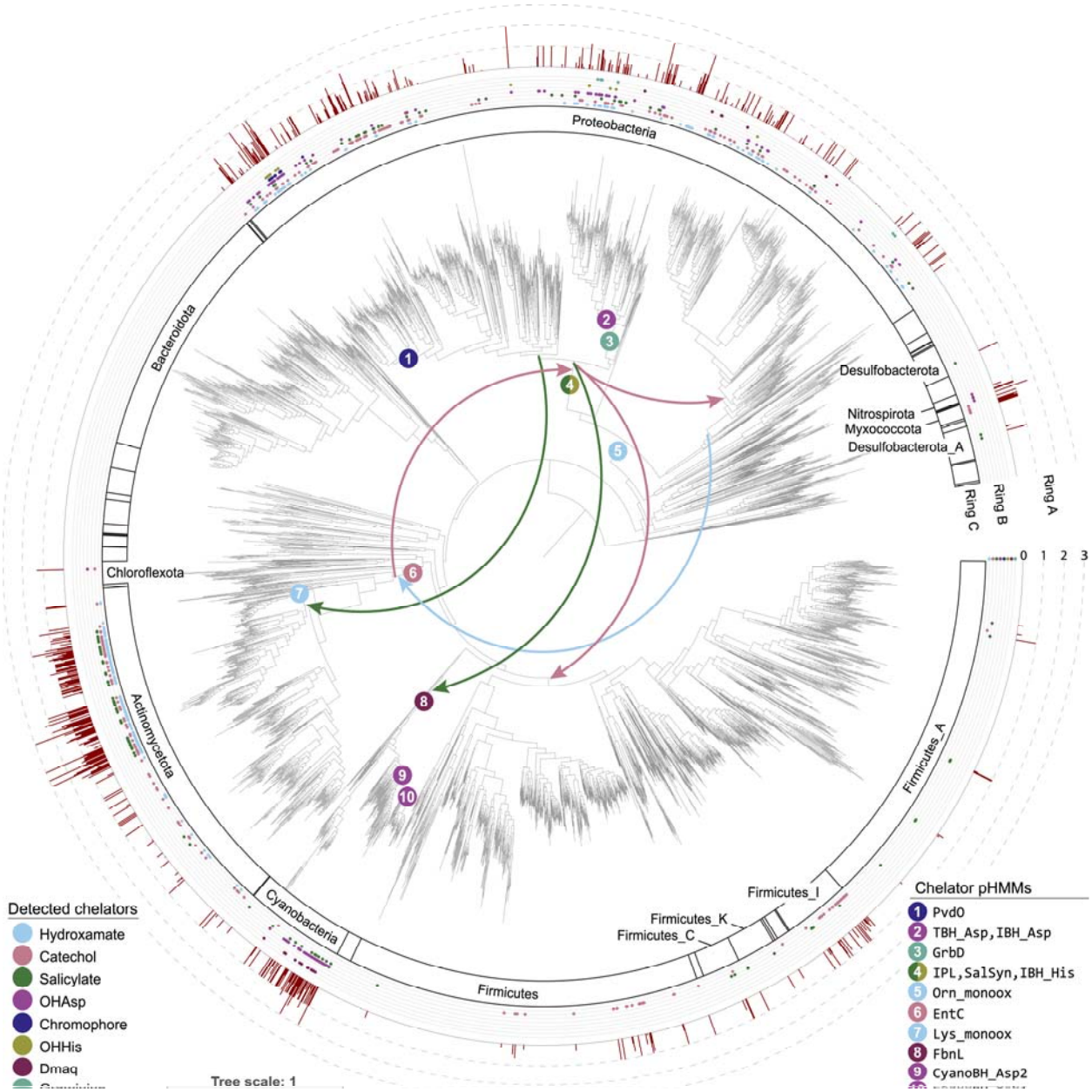


Figure 5. NRP metallophore biosynthesis across the bacterial kingdom. **Center:** The Genome Taxonomy Database (GTDB) phylogenetic tree (version r207), with strains collapsed to the REDgroup level.[46] Numbered circles indicate the most parsimonious origins of chelator pathways, as determined by reconciliation with eMPRess.[47] The bottom-right legend lists the specific hidden Markov models (pHMMs) associated with each estimated origin. Arrows indicate ancient horizontal gene transfers predicted by eMPRess. **Ring A:** Phylal divisions. Phyla with detected chelating groups are labeled using nomenclature from GTDB r207. **Ring B:** Chelator biosynthetic pathways detected in at least one member of each REDgroup. **Ring C:** Average number of detected NRP metallophore BGC regions per genome for each REDgroup. Annotations were mapped to the phylogenetic tree using iTOL v6.[50]

Discussion

Trace metal starvation shapes interactions within microbial communities and between bacteria and the host; therefore, natural and synthetic microbiomes cannot be understood without knowing the metallophore biosynthetic potential of the community. High-throughput biotechnological applications will benefit from *in silico* metallophore prediction due to the prohibitively high cost of isolation and characterization. To date, distinguishing peptidic metallophore BGCs from other NRPS BGCs has been largely limited to manual expert analyses, leading to blind spots in our understanding of microbes and their communities. We have now automated bacterial NRP metallophore prediction by extending the secondary metabolite prediction platform antiSMASH to detect genes involved in the biosynthesis of metal chelating moieties, enabling the first global analysis of bacterial metallophore biosynthetic diversity.

The presence of genes encoding catechols, hydroxamates, and other chelating groups is one of the most frequently used markers of a metallophore BGC.[9] We have formalized and automated the identification of eight chelator pathways, allowing us to detect 78% of NRP metallophore BGCs with a 3% false positive rate against a manually annotated set of NRPS clusters. Biosynthetic genes are detected with custom pHMMs and significance score cutoffs calibrated for accurate metallophore discovery, diminishing the ambiguity of interpreting gene annotations, protein families, and BLAST results. We acknowledge that human biases may have influenced which clusters were coded as putative metallophores during both algorithm development and testing; however, expert manual curation remains the most reliable method for NRP metallophore BGC detection. Unfortunately, 22% of manually identifiable metallophore BGCs could not be automatically distinguished from other NRPS clusters, as the algorithm developed (for the purpose of being easily integrated into antiSMASH) relies on the presence of one or more known chelator biosynthesis genes colocalized with the NRPS genes. Detection rates were also lower for fragmented genomes, a limitation (inherent to antiSMASH itself [37]) that hinders the identification of metallophore biosynthesis in metagenomes. As long-read sequencing of metagenomes becomes more common, we expect that detection will improve.

Recently, Crits-Christoph et al. demonstrated the use of transporter families to predict that a BGC encodes siderophore (or metallophore) biosynthesis.[33] Among our test dataset, the biosynthesis-based antiSMASH rule outperformed the transporter-based approach (F1 = 0.86 *versus* F1 = 0.69). However, some putative metallophore BGCs were only found using the transporter-based approach, and a combined either/or ensemble approach slightly outperformed the antiSMASH rule alone (F1 = 0.90). Biosynthetic- and transporter-based

techniques are thus complementary, and future work could incorporate transporter genes into antiSMASH metallophore prediction. We note that the reported transporter-based approach uses just three pHMMs from Pfam, while our biosynthetic detection requires many custom pHMMs. An extended set of metallophore-specific transporter pHMMs designed according to the same principles as those followed for the biosynthesis-related pHMMs could significantly improve detection by reducing false positives and capturing other families of transporters. The NRP metallophore BGCs discovered in this study could serve as a dataset for developing a more comprehensive model for metallophore transporter detection.

The diverse enzyme families responsible for the biosynthesis of NRP metallophore chelating groups (Figure 1B) evince that metal chelation has evolved multiple times, and we expect that more NRPS chelator substructures remain undiscovered. In fact, during manuscript preparation, the novel chelator 5-aminosalicylate was reported in the structure of the *Pseudonocardia* NRP siderophore pseudonochelin,[30] and we found several unexplored clades of Fe(II)/α-ketoglutarate- dependent amino acid β-hydroxylases that are likely involved in metallophore biosyntheses (Figure S2). Additionally, we have likely identified a new biosynthetic pathway in the genome of *Sporomusa termitida* DSM 4440, which encodes a partial menaquinone pathway in place of a salicylate synthase to putatively produce a novel karamomycin-like metallophore (Figure S3).[51] The modular nature of the pHMM-based detection rule will allow for new chelating groups to be added as their biosyntheses are experimentally characterized.

Metallophore BGC regions from representative genomes were compared to reference BGCs and de-replicated into gene cluster families (GCFs) with BiG-SCAPE (Figure 3). We found 1093 metallophore BGC regions that were dissimilar from any reference BGC, and almost 500 distinct BGC regions were found in only a single strain. Although significant biosynthetic diversity remains undiscovered, cluster de-replication will become increasingly important to avoid re-isolating known compounds. We also assessed the taxonomic distribution of NRP-metallophore BGC regions by mapping their presence onto a GTDB REDgroup phylogeny. We found that Cyanobacteria and Myxococcota were underrepresented in our analyses due to a relatively low number of published genomes. Considering that only a handful of NRP metallophores have been isolated from these two phyla, we suggest that Cyanobacteria and Myxococcota deserve coordinated efforts of genomic sequencing and experimental work to further characterize their metallophore diversity.

Finally, we used our dataset of detected BGCs and paired taxonomic data from GTDB to investigate the complex evolutionary history of chelating group biosynthesis by reconstructing

the most likely origin and major HGT events for each pathway with eMPRess (Fig. 5).[47] Catechols, hydroxamates, and salicylates were among the most widespread and ancient chelators with evidence of HGT between phyla. This widespread distribution suggests significant ecological relevance for these chelators in diverse bacterial lineages. Intriguingly, our timeline estimates place the origin of 2,3-DHB and Dmaq prior to the Great Oxygenation Event (~2.4 - 2.1 Ga), during an era of abundant, soluble ferrous iron. This result leads credence to the hypothesis that chelating molecules first evolved as metal detoxification mechanisms and were repurposed when oxidized iron became scarce.[3] Tracing ancient evolutionary events, particularly those involving multiple gene gains and losses, remains challenging due to the exponential increase in complexity as the number of possible events grows. More detailed examinations dedicated to each individual chelating group may yield deeper insights into the complex evolutionary history of these pathways. For example, the origin of hydroxamates must consider the homologous enzymes in NRPS-independent siderophore pathways, and we cannot yet state if metallophore-specific β-OHAsp biosynthesis is polyphyletic due to repeated incorporation into metallophores or a single incorporation followed by repeated transfer into non-chelating roles. Nevertheless, this study represents, to our knowledge, the first attempt at a large-scale phylogenetic analysis into the origin of chelating groups in bacteria.

By integrating chelator detection into antiSMASH, we have taken a major step towards accurate, automated NRP metallophore BGC detection. The new strategy affords a clear practical improvement over manual curation, and has already allowed for the high-throughput identification of thousands of likely NRP metallophore BGC regions, both in this study and in several other recently published analyses that have been enabled by early availability of our methodology.[52,53] A future antiSMASH module might predict metallophore activity more accurately with a machine learning algorithm that considers multiple forms of genomic evidence, including the presence of transporter genes, NRPS domain architecture and sequence, metal-responsive regulatory elements, and other markers of metallophore biosynthesis that are still limited to manual inspection.[9] In particular, regulatory elements will likely be required to accurately distinguish siderophores, zincophores, and other classes of metallophores. We also envision a later version that predicts specific structural features of the metallophore by identifying accessory biosynthetic genes. Implementation of NRP metallophore BGC detection into antiSMASH will enable scientists of diverse expertises to identify and quantify NRP metallophore biosynthetic pathways in their bacterial genomes of interest and promote large-scale investigations into the chemistry and biology of metallophores.

Methods

For all software, default parameters were used unless otherwise specified. All python, R, and bash scripts used in this paper, as well as underlying data, is available in the Supplemental Dataset, published to Zenodo: 10.5281/zenodo.17806971.

Profile hidden Markov model construction

Profile hidden Markov models (pHMMs) were built from biosynthetic genes in known metallophore pathways, supplemented with putative BGC genes where required (Figure S1 and Supplemental Dataset, `1_development/`). Amino acid sequences were aligned with MUSCLE (v3.8)[54] and pHMMs were constructed using hmmbuild (HMMER3).[55] pHMMs were tested against the MIBiG database (v2.0)[56] and an additional 37 NRP siderophore BGCs from literature (Supplemental Table 1) using hmmsearch (HMMER3). Rough bitscore significance cutoffs were determined for each pHMM. More precise cutoffs were assigned by testing against 28,688 NRPS BGC regions from the antiSMASH database (v3).[57] BGC regions containing genes near the rough cutoff were manually inspected to determine if these were likely metallophore BGCs. If no clear bitscore cutoff could be discerned, representative low-scoring putative true hits were added to the pHMM seed alignment. This process was repeated until a precise bitscore cutoff could be determined.

Phylogenetic analysis of Asp and His β-hydroxylases

Adequately-performing pHMMs for Asp and His β-hydroxylase subtypes could not be constructed using the above method. Siderophore β-hydroxylase functional subtypes were previously shown to form distinct phylogenetic clades.[22] An expanded phylogenetic analysis was performed to serve as a guide for pHMM construction (Supplemental Dataset, `1_development/hydroxylase_tree/`). NRPS BGC regions from the antiSMASH database (v3) were scanned for matches to previously reported β-hydroxylase pHMMs[22] and Pfam pHMMs for siderophore-related transporters (PF00593, PF01032, and PF01497[33,58]) using a modified version of antiSMASH v6.0. β-Hydroxylase genes meeting a relaxed bitscore cutoff of 300 (1070 total) were dereplicated with CD-HIT web server[59] and a sequence identity cutoff of 70%, giving 425 representative amino acid sequences. A multiple sequence alignment was created using hmmalign (HMMER3) and the TauD Pfam (PF02668),[58] and a maximum-likelihood phylogenetic tree was reconstructed with IQ-TREE (multicore v2.2.0-beta)[60] using the WAG+F+I+G4 evolutionary model. The presence of nearby transporters was mapped onto the phylogenetic

tree to identify clades or paraphyletic groups putatively involved in siderophore biosynthesis. Sequences in groups corresponding to previously reported TBH_Asp, IBH_Asp, and IBH_His subtypes and the novel putative CyanoBH_Asp1 and CyanoBH_Asp2 subtypes were extracted, and pHMMs were constructed and tested as described above.

Incorporation into antiSMASH

The pHMMs and cutoffs were added to antiSMASH as a single detection rule called "NRP-metallophore" with the following logic:

```
VibH_like or Cy_tandem or
(cds(Condensation and AMP-binding) and (
      (IBH_Asp and not SBH_Asp) or IBH_His or TBH_Asp or
      CyanoBH_Asp1 or CyanoBH_Asp2 or
      IPL or SalSyn or (EntA and EntC) or
      (GrbD and GrbE) or (FbnL and FbnM) or PvdO or PvdP or
      (Orn_monoox and not (KtzT or MetRS-like))
      Lys_monoox or VbsL))
```

Manual validation

A subset of RefSeq representative bacterial genomes was generated by randomly selecting one genome for each of the 330 genera determined by GTDB (Supplemental Dataset, `3_manual_testing/`). All NRPS BGC regions in the genomes were annotated with antiSMASH v6.1, yielding 758 BGC regions in the final testing dataset (Supplemental Table 2). The antiSMASH output for each BGC was manually inspected for evidence of NRP metallophore production, including genes encoding transporters, iron reductases, chelator biosynthesis, and known metallophore NRPS domain motifs. The same 758 BGC regions were classified as NRP metallophores using the chelator-based strategy described above. We compared the performance against the transporter-based strategy of Crits-Christoph *et al.*[33] Genes matching Pfam pHMMs for siderophore-related transporters (PF00593, PF01032, and PF01497)[33,58] were identified using a custom version of antiSMASH v6.1, in which the three pHMMs were added to the list of domains that are identified and labeled as "biosynthetic-additional". BGC regions were classified as metallophores if two of the three transporter families were present.[33] Each putative false positive was re-investigated before performance statistics were calculated, resulting in the reannotation of four BGCs.

BIG-SCAPE clustering

The 3,264 NRP metallophore BGC regions from RefSeq representative genomes (Supplemental Dataset, `2_refseq_reps_results/metallophores_Jun25.tar.gz`) were filtered to remove BGC regions on contig boundaries. The resulting 2,523 BGC regions, as well as 78 previously reported BGCs (Supplemental Table 2) were clustered using BiG-SCAPE v1.1.2 with the following settings: "`--no_classify --mix --cutoffs 0.3 0.4 0.5 --clans-off`". The network (Supplemental Dataset, `6_bigscape/mix_c0.50.network`) was imported to Cytoscape for figure preparation.

Phylogenetic mapping

Genome mining was performed on 62,291 GTDB representative genomes (59,851 after filtering; version r207)[44] using AntiSMASH v7.0beta,[13,44] with the inclusion of the NRP metallophore detection module. The outputs were analysed to identify predicted NRP-metallophore producers and categorized into distinct chelator groups based on predefined detection criteria. A total of 5,366 NRP-metallophores were identified, representing approximately 14% of all detected NRPS regions. To map the distribution patterns of these producers, the results were integrated with the GTDB tree. Due to the size of the tree, visualization tools such as iTOL[50] were impractical, prompting dereplication to a higher taxonomic rank. The GTDB tree was collapsed to the REDgroup level—a phylogenetically defined rank analogous to genera—allowing normalization to reflect the average number of NRP-metallophore biosynthetic gene clusters (BGCs) per genome within each REDgroup.[46]

To uncover the evolutionary history of siderophore biosynthesis, phylogenetic analyses and reconciliation were performed. Gene sequences for each chelator group were extracted from 4,060 complete BGCs, filtered to exclude BGC regions on contig boundaries, and clustered into Gene Cluster Families (GCFs) using BiG-SCAPE[61] with a 0.5 cutoff. From each GCF, one representative BGC was selected, resulting in a dataset of 1,108 clusters. Multiple sequence alignments (MSAs) were conducted using MAFFT v7,[62] and phylogenetic trees were constructed using FastTree 2 with the WAG model.[63] Evolutionary events, including gene duplication, loss, and horizontal gene transfer, were identified using phylogenetic reconciliation in eMPRess[47] by comparing gene trees to species trees. To validate the robustness of the inferred evolutionary scenarios, SalSyn and CyanoBH_Asp1 were also reconciled using the likelihood-based tool AleRax.[48] Reconciliation results were annotated using iTOL v6[50] for

visualization, manually mapping key evolutionary events onto the GTDB tree. Individual gene tree reconciliations are available in the Supplementary Dataset.

Data availability statement

All python, R, and bash scripts used in this paper, as well as underlying data, is available in the Supplemental Dataset, published to Zenodo: 10.5281/zenodo.17806971. The enterobactin, marinobactin, and ornicorrugatin BGCs have been submitted to the MIBiG repository with accession numbers BGC0003172, BGC0003173, BGC0003174, respectively.

Conflict of interest statement

The authors declare the following financial interests/personal relationships that may be considered as potential competing interests: M.H.M. is a member of the Scientific Advisory Board of Hexagon Bio.

Acknowledgements

This project has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (Starting Grant 948770-DECIPHER; ZR and MM), as well as from the US National Science Foundation (CHE-2108596; AB). BT and NZ were supported by H2020□FNR□11□2020: SECRETED—Grant agreement: 101000794. NZ was supported by the German Center for Infection Research TTU09.717. This work was supported by the Office of Navy Research Award Number N00014-23-2197. We thank Allegra Aron for providing useful feedback on the manuscript.

References

1. Hider, R. C. & Kong, X. Chemistry and biology of siderophores. *Nat. Prod. Rep.* **27**, 637–657 (2010).
2. Kraemer, S. M., Duckworth, O. W., Harrington, J. M. & Schenkeveld, W. D. C. Metallophores and Trace Metal Biogeochemistry. *Aquat. Geochem.* **21**, 159–195 (2015).
3. Kramer, J., Özkaya, Ö. & Kümmerli, R. Bacterial siderophores in community and host interactions. *Nat. Rev. Microbiol.* **18**, 152–163 (2020).
4. Soares, E. V. Perspective on the biotechnological production of bacterial siderophores and

their use. *Appl. Microbiol. Biotechnol.* **106**, 3985–4004 (2022).

5. Gu, S. *et al.* Competition for iron drives phytopathogen control by natural rhizosphere microbiomes. *Nat Microbiol* **5**, 1002–1010 (2020).
6. Gu, S. *et al.* Siderophore-Mediated Interactions Determine the Disease Suppressiveness of Microbial Consortia. *mSystems* **5**, e00811–19 (2020).
7. Behnsen, J. *et al.* Siderophore-mediated zinc acquisition enhances enterobacterial colonization of the inflamed gut. *Nat. Commun.* **12**, 7016 (2021).
8. Mehdiratta, K. *et al.* Kupyaphores are zinc homeostatic metallophores required for colonization of Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. U. S. A.* **119**, (2022).
9. Reitz, Z. L. & Medema, M. H. Genome mining strategies for metallophore discovery. *Curr. Opin. Biotechnol.* **77**, 102757 (2022).
10. Cézard, C., Farvacques, N. & Sonnet, P. Chemistry and biology of pyoverdines, Pseudomonas primary siderophores. *Curr. Med. Chem.* **22**, 165–186 (2015).
11. Süssmuth, R. D. & Mainz, A. Nonribosomal Peptide Synthesis-Principles and Prospects. *Angew. Chem. Int. Ed Engl.* **56**, 3770–3821 (2017).
12. Blin, K. *et al.* antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **49**, W29–W35 (2021).
13. Blin, K. *et al.* antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.* **51**, W46–W50 (2023).
14. Raymond, K. N., Dertz, E. A. & Kim, S. S. Enterobactin: an archetype for microbial iron transport. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3584–3588 (2003).
15. Serino, L. *et al.* Structural genes for salicylate biosynthesis from chorismate in Pseudomonas aeruginosa. *Mol. Gen. Genet.* **249**, 217–228 (1995).
16. Pelludat, C., Brem, D. & Heesemann, J. Irp9, encoded by the high-pathogenicity island of Yersinia enterocolitica, is able to convert chorismate into salicylate, the precursor of the siderophore yersiniabactin. *J. Bacteriol.* **185**, 5648–5653 (2003).

17. Keating, T. A., Marshall, C. G., Walsh, C. T. & Keating, A. E. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat. Struct. Biol.* **9**, 522–526 (2002).
18. Reitz, Z. L. & Butler, A. Precursor-directed biosynthesis of catechol compounds in Acinetobacter bouvetii DSM 14964. *Chem. Commun.* **56**, 12222–12225 (2020).
19. Bloudoff, K., Fage, C. D., Marahiel, M. A. & Schmeing, T. M. Structural and mutational analysis of the nonribosomal peptide synthetase heterocyclization domain provides insight into catalysis. *Proceedings of the National Academy of Sciences* **114**, 95–100 (2017).
20. Olucha, J. & Lamb, A. L. Mechanistic and structural studies of the N-hydroxylating flavoprotein monooxygenases. *Bioorg. Chem.* **39**, 171–177 (2011).
21. Heemstra, J. R., Walsh, C. T. & Sattely, E. S. Enzymatic Tailoring of Ornithine in the Biosynthesis of the Rhizobium Cyclic Trihydroxamate Siderophore Vicibactin. *J. Am. Chem. Soc.* **131**, 15317–15329 (2009).
22. Reitz, Z. L., Hardy, C. D., Suk, J., Bouvet, J. & Butler, A. Genomic analysis of siderophore β-hydroxylases reveals divergent stereocontrol and expands the condensation domain family. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 19805–19814 (2019).
23. Galica, T. *et al.* Cyanochelins, an Overlooked Class of Widely Distributed Cyanobacterial Siderophores, Discovered by Silent Gene Cluster Awakening. *Appl. Environ. Microbiol.* **87**, e0312820 (2021).
24. Hermenau, R. *et al.* Genomics-Driven Discovery of NO-Donating Diazeniumdiolate Siderophores in Diverse Plant-Associated Bacteria. *Angew. Chem. Int. Ed Engl.* **58**, 13024–13029 (2019).
25. Makris, C., Carmichael, J. R., Zhou, H. & Butler, A. C-Diazeniumdiolate Graminine in the Siderophore Gramibactin Is Photoreactive and Originates from Arginine. *ACS Chem. Biol.* **17**, 3140–3147 (2022).
26. Vinnik, V. *et al.* Structural and Biosynthetic Analysis of the Fabrubactins, Unusual

Siderophores from Agrobacterium fabrum Strain C58. *ACS Chem. Biol.* **16**, 125–135 (2021).

27. Nadal-Jimenez, P. *et al.* PvdP is a tyrosinase that drives maturation of the pyoverdine chromophore in Pseudomonas aeruginosa. *J. Bacteriol.* **196**, 2681–2690 (2014).
28. Ringel, M. T., Dräger, G. & Brüser, T. PvdO is required for the oxidation of dihydropyoverdine as the last step of fluorophore formation in Pseudomonas fluorescens. *J. Biol. Chem.* **293**, 2330–2341 (2018).
29. Kage, H., Kreutzer, M. F., Wackler, B., Hoffmeister, D. & Nett, M. An iterative type I polyketide synthase initiates the biosynthesis of the antimycoplasma agent micacocidin. *Chem. Biol.* **20**, 764–771 (2013).
30. Zhang, F. *et al.* Genome Mining and Metabolomics Unveil Pseudonochelin: A Siderophore Containing 5-Aminosalicylate from a Marine-Derived Pseudonocardia sp. Bacterium. *Org. Lett.* **24**, 3998–4002 (2022).
31. Wu, Q. *et al.* Metabolomics and Genomics Enable the Discovery of a New Class of Nonribosomal Peptidic Metallophores from a Marine Micromonospora. *J. Am. Chem. Soc.* **145**, 58–69 (2023).
32. Heather, Z. *et al.* A novel streptococcal integrative conjugative element involved in iron acquisition. *Mol. Microbiol.* **70**, 1274–1292 (2008).
33. Crits-Christoph, A., Bhattacharya, N., Olm, M. R., Song, Y. S. & Banfield, J. F. Transporter genes in biosynthetic gene clusters predict metabolite characteristics and siderophore activity. *Genome Res.* **31**, 239–250 (2020).
34. Jaremko, M. J., Davis, T. D., Corpuz, J. C. & Burkart, M. D. Type II non-ribosomal peptide synthetase proteins: structure, mechanism, and protein-protein interactions. *Nat. Prod. Rep.* **37**, 355–379 (2020).
35. Mihara, K. *et al.* Identification and transcriptional organization of a gene cluster involved in biosynthesis and transport of acinetobactin, a siderophore produced by Acinetobacter

baumannii ATCC 19606T. *Microbiology* **150**, 2587–2597 (2004).

36. González Carreró, M. I., Sangari, F. J., Agüero, J. & García Lobo, J. M. Brucella abortus strain 2308 produces brucebactin, a highly efficient catecholic siderophore. *Microbiology* **148**, 353–360 (2002).

37. Blin, K., Medema, M. H., Kottmann, R., Lee, S. Y. & Weber, T. The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters. *Nucleic Acids Res.* **45**, D555–D559 (2017).

38. Gross, H. & Loper, J. E. Genomics of secondary metabolite production by Pseudomonas spp. *Nat. Prod. Rep.* **26**, 1408–1446 (2009).

39. Matthijs, S., Budzikiewicz, H., Schäfer, M., Wathelet, B. & Cornelis, P. Ornicorrugatin, a New Siderophore from Pseudomonas fluorescens AF76. *Zeitschrift für Naturforschung C* **63**, (2008).

40. Ur□a Fern□ndez, D., Geoffroy, V., Sch□fer, M., Meyer, J.-M. & Budzikiewicz, H. Bacterial constituents CXIII structure revision of several pyoverdins produced by plant-growth promoting and plant-deleterious Pseudomonas species. *Monatsh. Chem.* **134**, 1421–1431 (2003).

41. Matthijs, S. *et al.* Pyoverdine and histicorrugatin-mediated iron acquisition in Pseudomonas thivervalensis. *Biometals* (2016) doi:10.1007/s10534-016-9929-1.

42. Martinez, J. S. & Butler, A. Marine amphiphilic siderophores: marinobactin structure, uptake, and microbial partitioning. *J. Inorg. Biochem.* **101**, 1692–1698 (2007).

43. Winkelmann, G., Cansier, A., Beck, W. & Jung, G. HPLC separation of enterobactin and linear 2,3-dihydroxybenzoylserine derivatives: a study on mutants of Escherichia coli defective in regulation (fur), esterase (fes) and transport (fepA). *Biometals* **7**, 149–154 (1994).

44. Parks, D. H. *et al.* GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy.

Nucleic Acids Res. **50**, D785–D794 (2022).

45. Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).
46. Gavriilidou, A. *et al.* Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. *Nat. Microbiol.* **7**, 726–735 (2022).
47. Santichaivekin, S. *et al.* eMPRess: a systematic cophylogeny reconciliation tool. *Bioinformatics* **37**, 2481–2482 (2021).
48. Morel, B., Williams, T. A., Stamatakis, A. & Szöllősi, G. J. AleRax: a tool for gene and species tree co-estimation and reconciliation under a probabilistic model of gene duplication, transfer, and loss. *Bioinformatics* **40**, btae162 (2024).
49. Marin, J., Battistuzzi, F. U., Brown, A. C. & Hedges, S. B. The timetree of prokaryotes: New insights into their evolution and speciation. *Mol. Biol. Evol.* **34**, 437–446 (2017).
50. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* **52**, W78–W82 (2024).
51. Shaaban, K. A. *et al.* Karamomycins A-C: 2-Naphthalen-2-yl-thiazoles from Nonomuraea endophytica. *J. Nat. Prod.* **82**, 870–877 (2019).
52. Mohite, O. S. *et al.* Pangenome mining of the*Streptomyces*genus redefines their biosynthetic potential. *bioRxiv* 2024.02.20.581055 (2024) doi:10.1101/2024.02.20.581055.
53. Jørgensen, T. S. *et al.* A treasure trove of 1034 actinomycete genomes. *Nucleic Acids Res.* **52**, 7487–7503 (2024).
54. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
55. Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
56. Kautsar, S. A. *et al.* MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Res.* **48**, D454–D458 (2020).
57. Blin, K., Shaw, S., Kautsar, S. A., Medema, M. H. & Weber, T. The antiSMASH database

version 3: increased taxonomic coverage and new query features for modular enzymes. *Nucleic Acids Res.* **49**, D639–D643 (2021).

58. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
59. Huang, Y., Niu, B., Gao, Y., Fu, L. & Li, W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **26**, 680–682 (2010).
60. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
61. Navarro-Muñoz, J. C. *et al.* A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* **16**, 60–68 (2020).
62. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
63. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).