Structures, Biosynthetic Pathways, and Biological Significance of Bacterial Aryl-Heterocycle Metallophores with Emphasis on Yersiniabactin-Type Derivatives

Martinus de Kruijff,[1] Sebastian Götze,[1] and Christine Beemelmanns[1,2]

[1] Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Department of Anti-Infectives from Microbiota, Helmholtz Centre for Infection Research (HZI), 66123 Saarbrücken, Germany

[2] Pharma Science Hub, Saarland University, 66123 Saarbrücken, Germany

Supporting Information

1. Methodology

1.1. Comparison of yersiniabactin-type metallophore biosynthetic gene clusters

Published biosynthetic gene clusters (BGCs) were extracted from the Minimum Information about a Biosynthetic Gene Cluster (MIBIG)[1] repository. For BGC entries missing in MIBIG, BGCs were manually extracted from genomes according to the respective publications. All BGCs were re-annotated using PROKKA (v1.14.6)[2], analyzed with antiSMASH (v7.1.0)[3], and manually curated to unify BGC boundaries. Gene cluster comparisons were performed with Biosynthetic Gene Similarity Clustering and Prospecting Engine (BiG-SCAPE, v2.0)[4] and cLINKER (v0.0.29)[5] using default settings.

1.2. Phylogenetic tree analyses

To compare the phylogeny of aryl-heterocycle producers, the 16S rRNA gene nucleotide sequence were obtained from published references, where available, or extracted manually from the genomes of the bacterial strains (Figure S1). Alignment of the sequences was performed with Muscle (v5.1)[6], trimmed using TrimAL (v1.5)[7] with the automated1 method, and the phylogenetic tree created with iqtree2 (v2.0.7)[8,9] specifying 1000 ultrafast bootstrap approximations (UFBoot)[10] and approximate likelihood ratio tests (aLRT). Phylogenetic trees were visualized employing R (v4.2.2, R core Team 2024, https://www.R-project.org) and the packages Ggtree[11] and Treeio[12].

The same procedure was performed for the creation of the phylogenetic tree for the C-A domains of the first NRPS module of BGCs of aryl-heterocycle producers. All amino acid sequences were manually truncated to include the C-A domains of the first NRPS module (Figure S2).

2. Supplementary Tables and Figures

Table S1A. List of known **aryl-heterocycle compound BGCs** with their corresponding bacterial producer, NCBI accession, and publication DOI. Only BGCs from publications with sufficient analytical data (HRMS and NMR) for the produced compound were used for further analysis. For complete list, see **Table S1B.**

Compound	Origin	NCBI_ACC	DOI
Acinetobactin	*Acinetobacter baumannii* ATCC19606	AB101202	10.1099/mic.0.27141-0
Agrobactin	*Agrobacterium tumefaciens* B6	LMVK01000001	10.1266/ggs.77.137
Amamistatin	*Nocardia altamirensis* DSM 44997	NZ_BDAY01000018	10.3762/bjoc.18.40
Amychelin	*Amycolatopsis* sp. AA4	NZ_CP024894	10.1021/ja203577e
Anachelin	*Anabaena cylindrica* PCC 7122	AP018166	10.1186/1471-2164-15-977
Anguibactin	*Vibrio anguillarum* 775	AY312585	10.1002/mbo3.65
Anthrochelin	*Luteibacter anthropi* DSM23190	JAARLZ010000002	10.1002/cbic.202300322
Attinimicin	*Pseudonocardia* sp. ICBG1293	NZ_JAHLJD000000000	10.1021/acscentsci.0c00978
Cahuitamycin	*Streptomyces gandocaensis* 12620-H2	KU363800	10.1038/ncomms10710
Cyanochelin	*Rivularia* sp. PCC 7116	CP003549	10.1128/AEM.03128-20
Ecteinamine	*Micromonospora* sp. WMMB482	JAHOPD010000020	10.1021/jacs.2c06410
Fimsbactin	*Acinetobacter baumannii* ATCC17978	NC_009085	10.1002/cbic.201200764
Frankobactin	*Parafrankia* sp. CH37	NZ_JADBID010000001	10.1021/acs.jnatprod.0c01291
Leptochelin	*Leptothoe* sp. ISB3NOV94-8A	JAXSQW000000000	10.1021/jacs.4c05399
Madurastatin	*Actinomadura* sp. WMMA1423	NZ_CP041244	10.1021/acs.orglett.9b02159
Massiliachelin	*Massilia* sp. NR 4-1	CP012201	10.3762/bjoc.15.128
Micacocidin	*Ralstonia solanacearum* GMI1000	AL646052	10.1128/AEM.05198-11
Mycobactin	*Mycobacterium tuberculosis* H37Rv	AL123456	10.1016/S1074-5521(98)90291-5
Nigribactin	*Vibrio nigripulchritudo* S2604	NZ_JXXT01000002	10.3390/md10112584
Nocardimicin	*Nocardia* sp. L-016	PQ621783	10.1038/ja.2005.77
Nocobactin	*Nocardia farcinica* IFM 10152	AP006618	10.1128/JB.00897-10
Parabactin	*Paracoccus denitrificans* PD1222	CP000489	10.1073/pnas.1401734111
Photobactin	*Photorhabdus luminescens* NS414	AY042783	10.1128/AEM.69.8.4706-4713.2003
Photoxenobactin	*Xenorhabdus szentirmaii* DSM 16338	NIBV00000000	10.1038/s41557-022-00923-2
Piscibactin	*Photobacterium damselae* subsp. *piscicida* DI21	KP100338	10.1128/AEM.01580-15
Pseudochelin A	*Pseudoalteromonas piscicida* EM138	JAXGFZ000000000	10.1021/acs.jnatprod.4c00376
Pseudomonine	*Pseudomonas fluorescens* AH2	EF484930	10.1128/JB.183.6.1909-1920.2001
Pseudonochelin	*Pseudonocardia* sp. WMMC193	JAKKDO010000003	10.1021/acs.orglett.2c01408
Pyochelin	*Burkholderia cepacia* ATCC 25416	CP012982	10.3389/fcimb.2017.00460
Serratiochelin	*Serratia plymuthica* V4	CP007439	10.1021/ja304941d
Sorangibactin	*Pendulispora rubella* MSr11367	OQ368584	10.1021/acschembio.3c00063
Terpenibactin	*Nocardia terpenica* IFM 0406	LWGR01000021	10.1002/cbic.202000062
Tetroazolemycin	*Streptomyces olivaceus* FXJ8.012	JAHSTE010000001	10.3390/md11051524
Ulbactin	*Brevibacillus brevis* NCTC2611	LR134338	10.1016/j.genrep.2016.10.006
Uniformide	*Nocardia uniformis* JCM3224	NZ_JABELX010000001	10.1021/acs.orglett.2c02092
Vibriobactin	*Vibrio cholerae* MBG14	U52150	10.1128/jb.179.22.7055-7062.1997
Vulnibactin	*Vibrio vulnificus* CMCP6	AE016796	10.3389/fmicb.2014.00001
Watasemycin	*Streptomyces venezuelae* ATCC 10712	FR845719	10.1039/c6sc03533g
Yersiniabactin	*Yersinia pestis* KIM6+	AF091251	10.1016/s1074-5521(98)90115-6

Table S2. List of included and excluded compounds of the YTM category including key chemical features and justifications.

Compound	Key feature	YTM	Justification
Acinetobactin	Histidine-type heterocycle	**N**	
Aerugine	-	**N**	Watasemycin shunt product
Aeruginoic acid	-	**N**	Pyochelin shunt product
Agrobactin	(nor)spermidine-type	**N**	
Amamistatin	C-terminal N–OH–cyclo-lysine	**N**	
Amychelin	Polypeptide chain	**N**	
Anachelin	Polypeptide chain	**N**	
Anguibactin	Histidine-type heterocycle	**N**	Lack of NRPS/PKS extender units between the heterocycles
Anthrochelin	2+ five-membered heterocycles connected with NRPS/PKS extenders	**Y**	
Attinimicin	Polypeptide chain	**N**	
Cahuitamycin	Polypeptide chain	**N**	
Carboxymycobactin	-	**N**	Mycobactin biosynthesis pathway
Cyanochelin	Polypeptide chain	**N**	
Dihydroaeruginoic Acid	-	**N**	Pyochelin shunt product
Ecteinamine	2+ five-membered heterocycles connected with NRPS/PKS extenders	**Y**	
Enantio-Pyochelin	-	**N**	Pyochelin shunt product
Escherichelin	-	**N**	Yersiniabactin shunt product
Fimsbactin	Amino acid linker	**N**	
Frankobactin	Polypeptide chain	**N**	
Leptochelin	2+ five-membered heterocycles connected with NRPS/PKS extenders	**Y**	
Madurastatin	Polypeptide chain	**N**	
Massiliachelin	2+ five-membered heterocycles connected with NRPS/PKS extenders	**Y**	
Micacocidin	2+ five-membered heterocycles connected with NRPS/PKS extenders	**Y**	
Mycobactin	C-terminal N–OH–cyclo-lysine	**N**	
Nigribactin	(nor)spermidine-type	**N**	
Nocardimicin	C-terminal N–OH–cyclo-lysine	**N**	
Nocobactin	C-terminal N–OH–cyclo-lysine	**N**	
Parabactin	(nor)spermidine-type	**N**	
Photobactin	Diamine linker	**N**	
Photoxenobactin	2+ five-membered heterocycles connected with NRPS/PKS extenders	**Y**	
Piscibactin	2+ five-membered heterocycles connected with NRPS/PKS extenders	**Y**	
Pre-Acinetobactin	Histidine-type heterocycle	**N**	Lack of NRPS/PKS extender units between the heterocycles
Pre-Pseudomonine	Histidine-type heterocycle	**N**	Lack of NRPS/PKS extender units between the heterocycles
Pseudochelin A	Amino acid linker	**N**	
Pseudomonine	Histidine-type heterocycle	**N**	
Pseudonochelin	Polypeptide chain	**N**	

Pyochelin	2+ five-membered heterocycles connected with NRPS/PKS extenders	**Y**	
Serratiochelin	Diamine linker	**N**	
Sorangibactin	2+ five-membered heterocycles connected with NRPS/PKS extenders	**Y**	
Terpenibactin	C-terminal N–OH–cyclo-lysine	**N**	
Tetroazolemycin	2+ five-membered heterocycles connected with NRPS/PKS extenders	**Y**	
Thiazostatin	-	**N**	Watasemycin shunt product
Ulbactin	2+ five-membered heterocycles connected with NRPS/PKS extenders	**Y**	
Uniformide	2+ five-membered heterocycles connected with NRPS/PKS extenders	**Y**	
Vibriobactin	(nor)spermidine-type	**N**	
Vulnibactin	(nor)spermidine-type	**N**	
Watasemycin	2+ five-membered heterocycles connected with NRPS/PKS extenders	**Y**	
Yersiniabactin	2+ five-membered heterocycles connected with NRPS/PKS extenders	**Y**	

Table S3. List of known, published aryl-heterocycle biosynthesis pathways (BGCs) with the respective publication DOI.

Compound	**DOI**
Acinetobactin	10.3389/fmicb.2021.752070
Amychelin	10.1021/ja203577e
Anguibactin	10.3389/fcimb.2017.00342
Anthrochelin	10.1002/cbic.202300322
Attinimicin	10.1021/acscentsci.0c00978
Cahuitamycin	10.1038/ncomms10710
Ecteinamine	10.1021/jacs.2c06410
Fimsbactin	10.1002/cbic.201200764
Frankobactin	10.1021/acs.jnatprod.0c01291
Leptochelin	10.1021/jacs.4c05399
Madurastatin	10.1021/acs.orglett.9b02159
Massiliachelin	10.3762/bjoc.19.69
Micacocidin	10.1016/j.chembiol.2013.04.010
Mycobactin	10.1021/acs.jmedchem.0c01176
Nocardimicin	10.1021/acs.jnatprod.4c01140
Nocobactin	10.1128/jb.00897-10
Photoxenobactin E	10.1021/acschembio.2c00367
Piscibactin	10.1002/ejoc.201200818
Pseudomonine	10.1039/B817075B
Pseudonochelin	10.1021/acs.orglett.2c01408
Pyochelin	10.1021/acs.biochem.4c00276
Serratiochelin	10.1021/ja304941d
Sorangibactin	10.1021/acschembio.3c00063
Watasemycin	10.1039/C6SC03533G
Yersiniabactin	10.1128/AEM.69.11.6698-6702.2003

Table S4. List of known, published aryl-heterocycle BGCs without biosynthesis pathways with the respective publication DOI.

Compound	DOI
Agrobactin	10.1266/ggs.77.137
Amamistatin	10.3762/bjoc.18.40
Anachelin	10.1186/1471-2164-15-977
Cyanochelin	10.1128/AEM.03128-20
Nigribactin	10.3390/md10112584
Parabactin	10.1073/pnas.1401734111
Pseudochelin A	10.1021/acs.orglett.2c01408
Terpenibactin	10.1002/cbic.202000062
Tetroazolemycin	10.3390/md11051524
Ulbactin	10.1016/j.genrep.2016.10.006
Vibriobactin	10.1128/jb.179.22.7055-7062.1997
Vulnibactin	10.3389/fmicb.2014.00001

Table S5. List of known aryl-heterocycle compounds for which neither BGC nor genomic data has been reported (until March 2025)

Compound	DOI
Aerugine	10.1039/c6sc03533g
Asterobactin	10.7164/antibiotics.55.593
BE-32030	10.7164/antibiotics.50.815
Brasilibactin	10.1021/np0496385
Karamomycin	10.1021/acs.jnatprod.8b00928
Pulicatin	10.1021/np100588c
Thiazostatin	10.1039/c6sc03533g
Transvalencin Z	10.1038/ja.2006.53

Table S6. List of known aryl-heterocycle producers with their corresponding bacterial origin used for 16S rRNA sequence extraction and their respective publication digital object identifiers (DOIs).

Compound	Origin	DOI
2-Hydroxybenzamide	*Streptomyces sp*. RM-14-6	10.1021/acs.jnatprod.6b00948
Acinetobactin	*Acinetobacter baumannii* ATCC19606	10.1007/BF00301846
Aerugine	*Pseudomonas aeruginosa* 590	10.1007/BF00597805
Agrobactin	*Agrobacterium tumefaciens* B6	10.1016/S0021-9258(17)37736-0
Amamistatin	*Nocardia altamirensis* DSM 44997	10.3762/bjoc.18.40
Amychelin	*Amycolatopsis sp*. AA4	10.1021/ja203577e
Anguibactin	*Vibrio anguillarum* 775	10.1128/jb.167.1.57-65.1986
Anguibactin	*Vibrio sp*. DS40M4	10.1021/np900750g
Anthrochelin	*Luteibacter anthropi* DSM23190	10.1002/cbic.202300322
Attinimicin	*Pseudonocardia sp*. ICBG1293	10.1021/acscentsci.0c00978
Brasilibactin	*Nocardia brasiliensis* IFM 0995	10.1021/np0496385
Cahuitamycin	*Streptomyces gandocaensis* 12620-H2	10.1038/ncomms10710
Carboxymycobactin	*Mycolicibacterium smegmatis* NCIMB 8548	10.1023/A:1009249122146
Cyanochelin	*Rivularia sp*. PCC 7116	10.1128/AEM.03128-20
Ecteinamine	*Micromonospora sp*. WMMB482	10.1021/jacs.2c06410
Enantio-Pyochelin	*Pseudomonas protegens* CHA0	10.1074/jbc.M707039200
Escherichelin	*Escherichia coli* UTI89	10.1172/JCI92464
Fimsbactin	*Acinetobacter baumannii* ATCC17978	10.1002/cbic.201200764
Fimsbactin	*Acinetobacter baylyi* ADP1	10.1002/cbic.201200764
Frankobactin	*Parafrankia sp*. CH37	10.1021/acs.jnatprod.0c01291
Karamomycin	*Nonomuraea sp*. GW58/450	10.1021/acs.jnatprod.8b00928
Leptochelin	*Leptothoe sp*. ISB3NOV94-8A	10.1021/jacs.4c05399
Madurastatin	*Actinomadura sp*. WMMA1423	10.1021/acs.orglett.9b02159
Massiliachelin	*Massilia sp*. NR 4-1	10.3762/bjoc.15.128
Micacocidin	*Ralstonia solanacearum* GMI1000	10.1128/AEM.05198-11
Nigribactin	*Vibrio nigripulchritudo* S2604	10.3390/md10112584
Nocardichelin	*Nocardia sp*. Acta 3026	10.1021/np060612i
Nocardimicin	*Nocardia nova* JCM 6044	10.1038/ja.2005.77
Nocobactin	*Nocardia farcinica* IFM 10152	10.1128/JB.00897-10
Oxachelin	*Streptomyces sp*. RM-14-6	10.1021/acs.jnatprod.6b00948
Parabactin	*Paracoccus dentrificans* ATCC19367	10.1016/S0021-9258(17)39542-X
Photoxenobactin	*Xenorhabdus szentirmaii* DSM 16338	10.1038/s41557-022-00923-2
Photoxenobactin	*Vibrio anguillarum* RV22	10.1021/acs.inorgchem.3c04076
Piscibactin	*Photobacterium damselae* subsp. *piscicida* DI21	10.1128/AEM.01580-15
Pseudochelin A	*Pseudoalteromonas piscicida* EM138	10.1021/acs.jnatprod.4c00376
Pseudomonine	*Pseudomonas fluorescens* WCS374	10.1128/JB.183.6.1909-1920.2001
Pseudonochelin	*Pseudonocardia sp*. WMMC193	10.1021/acs.orglett.2c01408
Pulicatin	*Streptomyces sp*. CP32	10.1021/np100588c
Pulicatin	*Pantoea agglomerans* Pa	10.3390/biom10020268
Pulicatin	*Streptomyces ardesiacus* NBU2966	10.1016/j.phytochem.2021.113078
Pyochelin	*Pseudomonas aeruginosa* PAO1	10.1128/jb.137.1.357-364.1979
Serratiochelin	*Serratia marcescens* TW	10.1515/znc-1994-1-203
Serratiochelin	*Serratia plymuthica* V4	10.1021/ja304941d
Sorangibactin	*Pendulispora rubella* MSr11367	10.1021/acschembio.3c00063
Spoxazomicin	*Streptosporangium oxazolinicum* K07-0460	10.1038/ja.2011.16
Spoxazomicin	*Streptomyces sp*. RM-14-6	10.1021/acs.jnatprod.6b00948
Spoxazomicin	*Streptomyces ardesiacus* NBU2966	10.1016/j.phytochem.2021.113078
Terpenibactin	*Nocardia terpenica* IFM 0406	10.1002/cbic.202000062
Tetroazolemycin	*Streptomyces olivaceus* FXJ8.012	10.3390/md11051524
Thiazostatin	*Actinoplanes sp*. SE50/110	10.3389/fbioe.2024.1497138
Thiazostatin	*Streptomyces ardesiacus* NBU2966	10.1016/j.phytochem.2021.113078
Ulbactin	*Brevibacillus brevis* NCTC2611	10.1021/acs.orglett.6b00531
Uniformide	*Nocardia uniformis* JCM3224	10.1021/acs.orglett.2c02092
Vibriobactin	*Vibrio cholerae* O395	10.1021/bi001651a
Yersiniabactin	*Yersinia pestis* KIM6+	10.1016/s1074-5521(98)90115-6
Yersiniabactin	*Escherichia coli* Nissle 1917	10.1038/s41467-021-27297-2

Table S7. Categorization of BGCs encoding for aryl-heterocycle natural products into GCFs. BGCs without a known gene symbol were marked with an asterisk (*) and an interim three letter code was assigned. Bold BGCs are hallmarks for referencing the given GCFs.

Mycobactin	***mbt***	**GCF1**
Amamistatin	*ama**	
Amychelin	*amc*	
Cahuitamycin	*cah*	
Frankobactin	*fra*	
Nocardimicin	*noc*	
Nocobactin	*nbt*	
Terpenibactin	*ter*	
Madurastatin	***mad***	**GCF2**
Attinimicin	*att*	
Pseudonochelin	*psn*	
Anachelin	***ana***	**GCF3**
Cyanochelin	*ccs*	
Leptochelin	*lec*	
Acinetobactin	***bas***	**GCF4**
Agrobactin	*agb*	
Anguibactin	*ang*	
Fimsbactin	*fbs*	
Nigribactin	*ngb**	
Parabactin	*pbc**	
Photobactin	*phb*	
Pseudochelin	*mxc*	
Pseudomonine	*pms*	
Serratiochelin	*sch*	
Vibriobactin	*vib*	
Vulnibactin	*vv2*	
Yersiniabactin	***ybt***	**GCF5**
Photoxenobactin	*pxb*	
Piscibactin	*pcb*	
Micacocidin	***mic***	**GCF6**
Massiliachelin	*mas**	
Pyochelin	***pch***	**GCF7**
Anthrochelin	*ant*	
Ecteinamine	*ect*	
Sorangibactin	*srb*	
Tetroazolemycin	*tet**	
Ulbactin	*ubt**	
Uniformide	*uni**	
Watasemycin	*sven*	

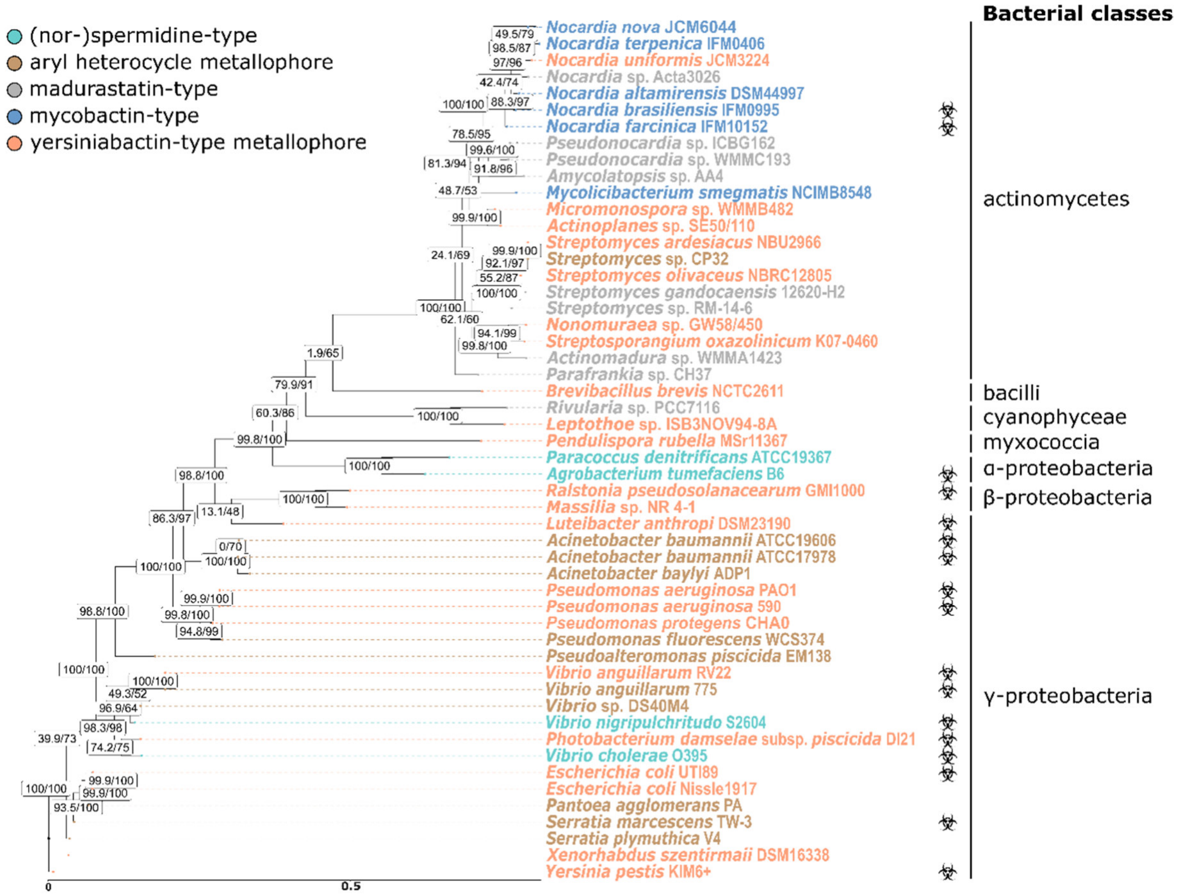

Figure S1. Phylogeny tree of 16S rRNA genes from aryl-heterocycle natural product producers with alRT and bootstrap values. Phylogenetic tree was created with nucleotide 16S rRNA gene sequences.[6–12] 1000 replicates for alRT and bootstrap were used for tree construction. The text colour indicates the produced natural product group and pathogenic strains were highlighted by a biohazard symbol. Vertical lines show affiliations to bacterial classes. The branch length displays substitutions per site and the tree is depicted as unrooted.

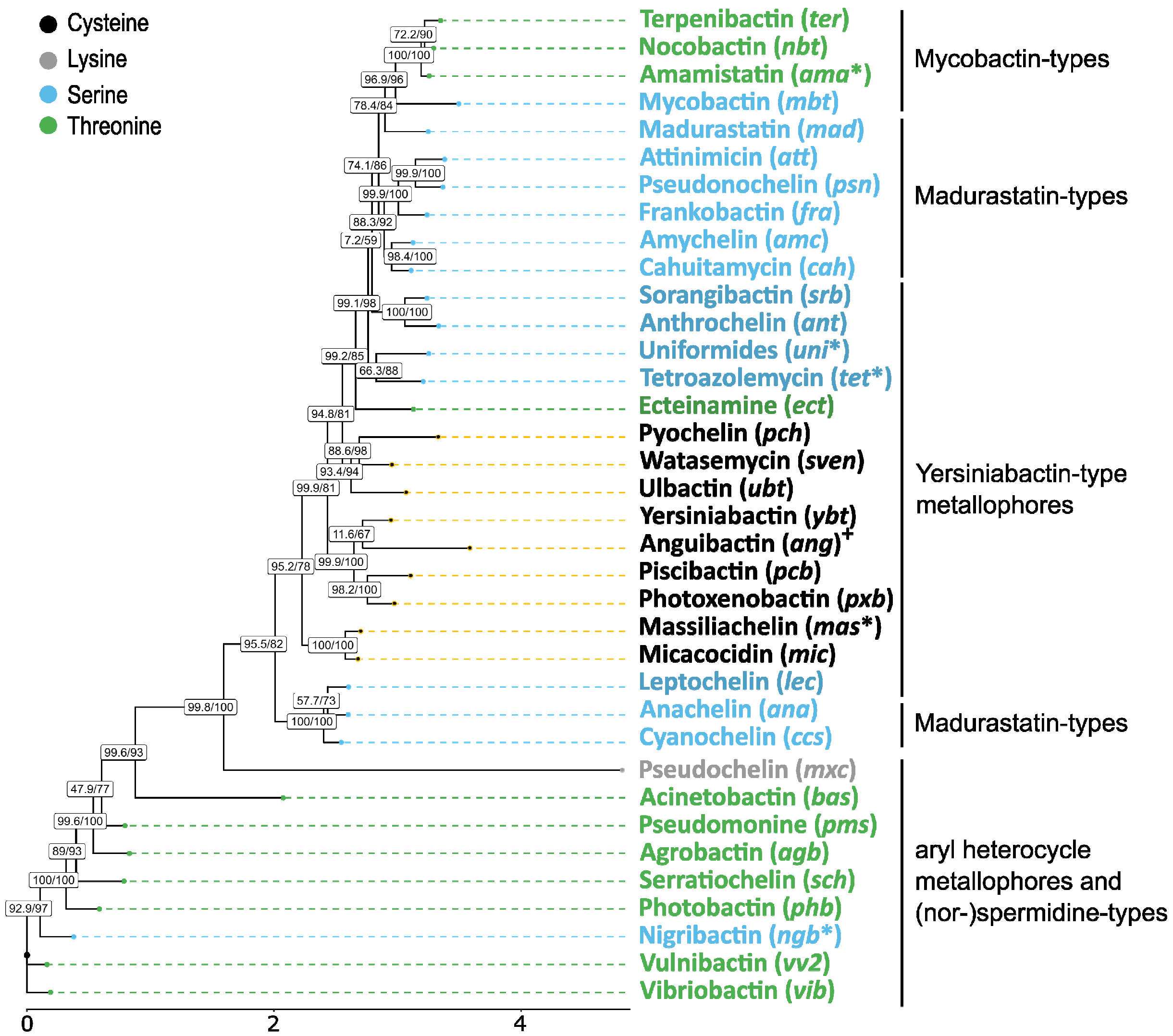

Figure S2. Phylogenetic tree of C-A domains of the first NRPS module of the aryl-heterocycle compounds with aIRT and bootstrap values. Phylogenetic tree was created with amino acid sequences.[6–12] 1000 replicates for aIRT and bootstrap were used for tree construction. The text colour of the natural products indicates the A domain specificity. Vertical lines indicate natural product structure categories. BGCs without a known gene symbol were marked with an asterisk (*) and an interim three letter code was assigned. Anguibactin (+) showed notably low bootstrap and aIRT values. The grey box highlights the YTMs. The branch length displays substitutions per site and the tree is depicted as unrooted.

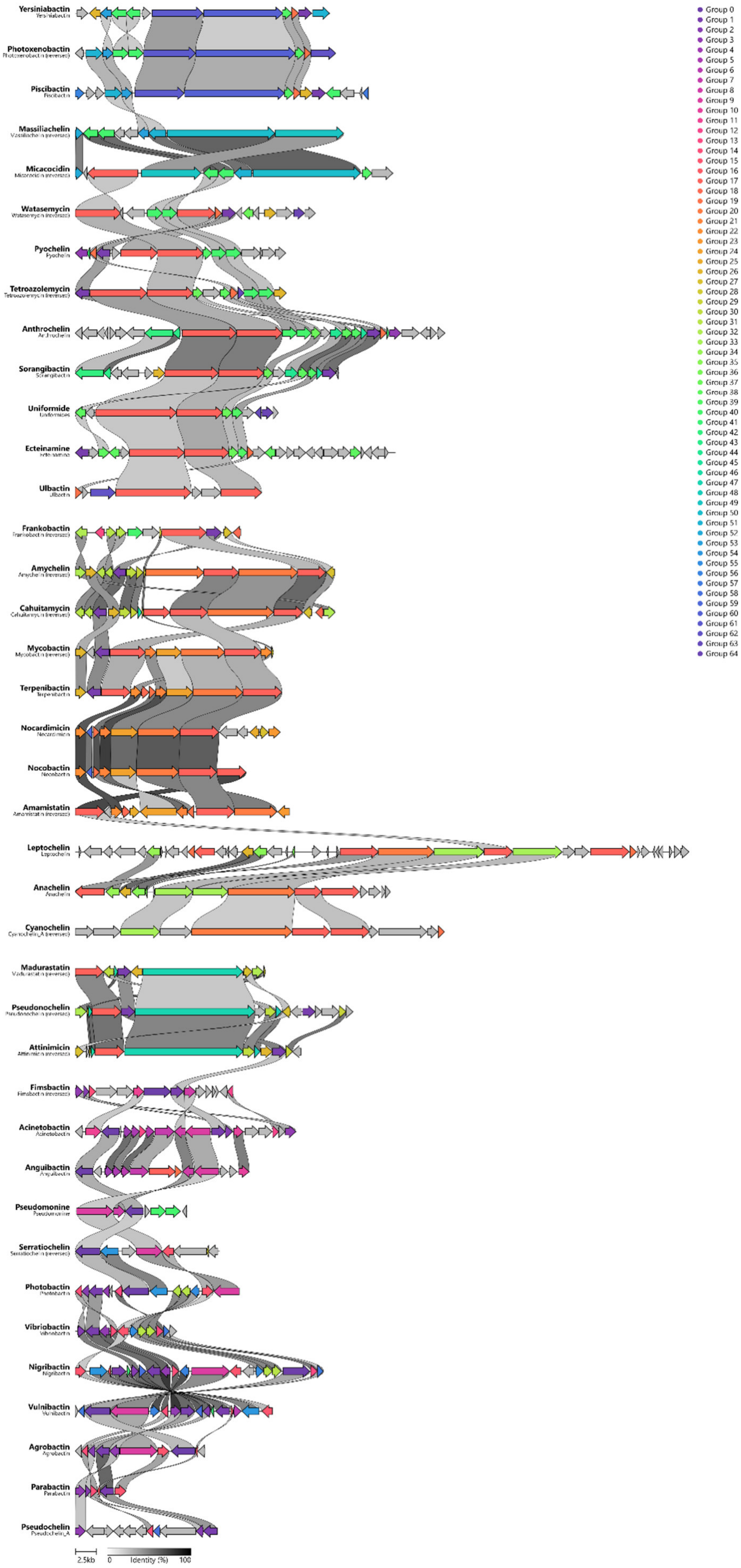

Figure S3. Full BGCs comparison

cLINKER comparison of selected 39 BGCs (Table S7) encoding for of aryl-heterocycle metallophores using the default identity cut-off of 30%.

4. References

1 B. R. Terlouw, K. Blin, J. C. Navarro-Muñoz, N. E. Avalon, M. G. Chevrette, S. Egbert, S. Lee, D. Meijer, M. J. J. Recchia, Z. L. Reitz, J. A. van Santen, N. Selem-Mojica, T. Tørring, L. Zaroubi, M. Alanjary, G. Aleti, C. Aguilar, S. A. A. Al-Salihi, H. E. Augustijn, J. A. Avelar-Rivas, L. A. Avitia-Domínguez, F. Barona-Gómez, J. Bernaldo-Agüero, V. A. Bielinski, F. Biermann, T. J. Booth, V. J. Carrion Bravo, R. Castelo-Branco, F. O. Chagas, P. Cruz-Morales, C. Du, K. R. Duncan, A. Gavriilidou, D. Gayrard, K. Gutiérrez-García, K. Haslinger, E. J. N. Helfrich, J. J. J. van der Hooft, A. P. Jati, E. Kalkreuter, N. Kalyvas, K. B. Kang, S. Kautsar, W. Kim, A. M. Kunjapur, Y.-X. Li, G.-M. Lin, C. Loureiro, J. J. R. Louwen, N. L. L. Louwen, G. Lund, J. Parra, B. Philmus, B. Pourmohsenin, L. J. U. Pronk, A. Rego, D. A. B. Rex, S. Robinson, L. R. Rosas-Becerra, E. T. Roxborough, M. A. Schorn, D. J. Scobie, K. S. Singh, N. Sokolova, X. Tang, D. Udwary, A. Vigneshwari, K. Vind, S. P. J. M. Vromans, V. Waschulin, S. E. Williams, J. M. Winter, T. E. Witte, H. Xie, D. Yang, J. Yu, M. Zdouc, Z. Zhong, J. Collemare, R. G. Linington, T. Weber and M. H. Medema, *Nucleic Acids Res.*, 2023, **51**, D603-D610.

2 T. Seemann, *Bioinformatics*, 2014, **30**, 2068–2069.

3 K. Blin, S. Shaw, H. E. Augustijn, Z. L. Reitz, F. Biermann, M. Alanjary, A. Fetter, B. R. Terlouw, W. W. Metcalf, E. J. N. Helfrich, G. P. van Wezel, M. H. Medema and T. Weber, *Nucleic Acids Res.*, 2023, **51**, W46-W50.

4 J. C. Navarro-Muñoz, N. Selem-Mojica, M. W. Mullowney, S. A. Kautsar, J. H. Tryon, E. I. Parkinson, E. L. C. de Los Santos, M. Yeong, P. Cruz-Morales, S. Abubucker, A. Roeters, W. Lokhorst, A. Fernandez-Guerra, L. T. D. Cappelini, A. W. Goering, R. J. Thomson, W. W. Metcalf, N. L. Kelleher, F. Barona-Gomez and M. H. Medema, *Nat. Chem. Biol.*, 2020, **16**, 60–68.

5 C. L. M. Gilchrist and Y.-H. Chooi, *Bioinformatics*, 2021, **37**, 2473–2475.

6 R. C. Edgar, *Nat. Commun.*, 2022, **13**, 6968.

7 S. Capella-Gutiérrez, J. M. Silla-Martínez and T. Gabaldón, *Bioinformatics*, 2009, **25**, 1972–1973.

8 B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler and R. Lanfear, *Mol. Biol. Evol.*, 2020, **37**, 1530–1534.

9 S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler and L. S. Jermiin, *Nat. Methods*, 2017, **14**, 587–589.

10 D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh and S. Le Vinh, *Mol. Biol. Evol.*, 2018, **35**, 518–522.

11 S. Xu, L. Li, X. Luo, M. Chen, W. Tang, L. Zhan, Z. Dai, T. T. Lam, Y. Guan and G. Yu, *iMeta*, 2022, **1**, e56.

12 L.-G. Wang, T. T.-Y. Lam, S. Xu, Z. Dai, L. Zhou, T. Feng, P. Guo, C. W. Dunn, B. R. Jones, T. Bradley, H. Zhu, Y. Guan, Y. Jiang and G. Yu, *Mol. Biol. Evol.*, 2020, **37**, 599–603.