ACTA
AMAZONICA

http://dx.doi.org/10.1590/1809-4392202304142

BIOTECHNOLOGY - ORIGINAL ARTICLE

Exploring the biotechnological potential of Amazonian microorganisms through *in silico* genome analysis for detection of biosynthetic gene clusters

Eric de Lima Silva MARQUES[1,2*], Adriana Barros de Cerqueira e SILVA[3], Mariana Barros de Cerqueira e SILVA[3]

[1] Universidade Federal de Alagoas, Instituto de Ciências Farmacêuticas, Av. Lourival Melo Mota, s/n, Tabuleiro do Martins, Maceió, Alagoas, Brazil

[2] Current affiliation: Instituto Leônidas e Maria Deane (Fiocruz Amazônia), Manaus, Amazonas, Brazil

[3] Universidade Federal do Amazonas, Faculdade de Ciências Farmacêuticas, Av. Jauary Marinho, Setor Sul, Coroado, Manaus, Amazonas, Brazil

*Corresponding author: marques.ericls@gmail.com

ABSTRACT

The discovery of novel bioactive compounds has significant implications for diverse biotechnological applications. However, microbial genomes remain largely underexplored regarding their potential to produce secondary metabolites. Here, we explore the biotechnological potential of microorganisms from the highly biodiverse Amazon region through *in silico* genome mining. A total of 40 bacterial genomes originating from water, soil, and animal and plant-associated microorganisms were selected from a public database and analyzed using AntiSMASH and PRISM, predicting, respectively, 402 and 195 biosynthetic gene clusters (BGCs) related to secondary metabolite production. The most frequent BGCs were associated with polyketides, nonribosomal peptides, and ribosomally synthesized and post-translationally modified peptides, diverse classes of secondary metabolites commonly associated with biotechnological applications. To evaluate similarity with previously characterized clusters, the predicted BGCs were compared to entries in the MIBiG repository using BiG-SCAPE, revealing 12 clusters with close relationships to known BGCs. The remaining clusters showed low similarity, indicating that many potentially novel biosynthetic pathways remain uncharacterized, highlighting the limitations of current reference databases and the need for experimental validation. The results highlight the underexplored biotechnological potential of Amazonian microorganisms and reinforce the importance of expanding microbial genome sequencing and mining efforts in this biodiversity hotspot. The findings have significant implications for the discovery of novel bioactive compounds with diverse applications.

KEYWORDS: AntiSMASH, NRP, polyketides, PRISM, RiPP, secondary metabolites

Explorando o potencial biotecnológico de microrganismos amazônicos por meio da análise *in silico* de genomas para detecção de agrupamentos de genes biossintéticos

RESUMO

A descoberta de novos compostos bioativos tem implicações significativas para diversas aplicações biotecnológicas. No entanto, os genomas microbianos permanecem amplamente subexplorados quanto ao seu potencial de produzir metabólitos secundários. Aqui, exploramos o potencial biotecnológico de microrganismos da altamente biodiversa região amazônica, por meio da mineração *in silico* do genoma. Um total de 40 genomas bacterianos, oriundos de microrganismos associados à água, solo, animais e plantas da Amazônia, foram selecionados a partir de uma base de dados pública e analisados com as ferramentas AntiSMASH e PRISM, que previram, respectivamente, 402 e 195 clusters de genes biossintéticos (BGCs) relacionados a produção de metabólitos secundários. Os BGCs mais frequentes foram associados a poliquetídeos, peptídeos não ribossômicos e peptídeos sintetizados ribossomicamente e modificados pós-traducionalmente, classes de metabólitos secundários comumente relacionadas a aplicações biotecnológicas. Para avaliar a similaridade com clusters previamente caracterizados, os BGCs preditos foram comparados às entradas do repositório MIBiG utilizando BiG-SCAPE, revelando 12 clusters com alta similaridade a BGCs conhecidos. Os demais clusters mostraram baixa similaridade, indicando que muitas vias biossintéticas potencialmente novas permanecem não caracterizadas, evidenciando as limitações das bases de dados de referência atuais e a necessidade de validação experimental. Em geral, os resultados ressaltam o potencial biotecnológico pouco explorado dos microrganismos amazônicos e reforçam a importância de expandir os esforços de sequenciamento e mineração de genomas microbianos nesse hotspot de biodiversidade. As descobertas têm implicações significativas para novos compostos bioativos com aplicações diversas.

PALAVRAS-CHAVE: AntiSMASH, NRP, poliquetideos, PRISM, RiPP, metabolitos secundários

CITE AS: Marques, E.L.S.; Silva, A.B.C.; Silva, M.B.C. 2026. Exploring the biotechnological potential of Amazonian microorganisms through *in silico* genome analysis for detection of biosynthetic gene clusters. *Acta Amazonica* 56: e56bt23414.

INTRODUCTION

The Amazon region, one of the most biodiverse areas on the planet, plays a pivotal role in maintaining global ecosystem stability (Ellwanger *et al*. 2020; da Cruz *et al*. 2021). Its extensive rainforest and intricate natural systems harbor many plants, animals, and microorganisms, many of which remain unexplored. These organisms represent a valuable reservoir for biotechnology and the discovery of novel bioactive compounds (Lima *et al*. 2015; Carvalho *et al*. 2016; Pereira *et al*. 2017).

The Amazon rainforest contains a wide range of ecological niches shaped by complex environmental gradients, including high humidity, temperature variation, acidic and nutrient-poor soils, seasonal flooding, and intense competition for resources (Levine *et al*. 2016; Wittmann *et al*. 2022). These conditions act as selective pressures that drive microbial adaptation and diversification, potentially leading to the evolution of novel metabolic pathways and biosynthetic capabilities. For example, microorganisms from floodplain forests must withstand rapid changes in oxygen availability and osmotic stress, selecting for specialized stress-response mechanisms and secondary metabolites (Wittmann *et al*. 2022). Despite this promising potential, the microbial diversity of the Amazon remains largely underexplored (Pereira *et al*. 2017), partly due to limited sampling efforts and insufficient targeted research.

Several studies have demonstrated the high microbial diversity in Amazonian environments, especially in preserved areas and/or in association with animals and plants (Fonseca *et al*. 2018; Kroeger *et al*. 2018; Thompson *et al*. 2023). The biotechnological potential of Amazonian microorganisms has been assessed with emphasis on agricultural applications (Silva *et al*. 2014; Martins da Costa *et al*. 2018; 2019; Cabral Michel *et al*. 2021; Lopes *et al*. 2025), enzyme production (Mendes *et al*. 2015; Almeida Lima *et al*. 2023), antimicrobial activity (Cerqueira dos Santos *et al*. 2024; Rodrigues *et al*. 2024), and bioremediation (Cardona *et al*. 2022; Corral-García *et al*. 2024). The genetic diversity of the Amazon River microbiome and its capacity to degrade complex organic compounds (Santos-Júnior *et al*. 2020) emphasize the vast biotechnological potential of Amazonian microorganisms. However, the general scarcity of studies reinforces the importance of investigating biosynthetic gene clusters (BGCs) in Amazonian microorganisms. BGCs are groups of physically co-located genes in microbial genomes that collectively encode the biosynthesis of secondary metabolites, which may include novel natural products with biotechnological potential (Bauman *et al*. 2021; Bhattacharjee *et al*. 2023).

Genome mining has emerged as a powerful approach to uncover the biotechnological potential of microorganisms by analyzing their genomic data (Bhattacharjee *et al*. 2023). The Amazon's rich microbial communities represent a vast reservoir of bioactive metabolites that can be identified through *in silico* approaches (Bauman *et al*. 2021). These metabolites include antibiotics, antifungals, antitumor agents, immunomodulators, enzyme inhibitors, antidiabetics, analgesics, vasodilators, among others, with potential applications in medicine, agriculture and industry (Bhattacharjee *et al*. 2023).

Unlike primary metabolites that support essential cellular processes, secondary metabolites are specialized compounds produced by bacteria to enhance survival under specific environmental conditions, such as competition, stress, or host interaction (Santamaria *et al*. 2022). Most commercially important secondary metabolites are derived from Actinomycetota (Bhattacharjee *et al*. 2023), although an increasing number of metabolites from non-Actinomycetota taxa have shown promising biotechnological potential (Sharrar *et al*. 2020; Wei *et al*. 2021).

Secondary metabolites are synthesized by BGCs, which can be identified with specialized bioinformatic tools (Blin *et al*. 2023; Bhattacharjee *et al*. 2023). Three major classes of BGC-related metabolites are polyketides (PKS), nonribosomal peptides (NRP), and ribosomally synthesized and post-translationally modified peptides (RiPP). PKS are structurally diverse compounds synthesized through the polymerization of extender units such as malonyl-CoA, methylmalonyl-CoA, and propionyl-CoA, which are further modified to yield molecules with antimicrobial, antitumor (Kormanec *et al*. 2020), and pesticidal properties (Li *et al*. 2021). NRPs are peptides assembled by nonribosomal peptide synthetases, often forming cyclic and/or branched structures, both standard and non-proteinogenic amino acids with diverse biological activities (McErlean *et al*. 2019; Duban *et al*. 2022). RiPPs are ribosomally synthesized peptides undergoing complex post-translational modifications that contribute to structural diversity and diverse biological functions (Hetrick and van der Donk 2017).

Given the high diversity of microorganisms in the Amazon and advances in next-generation sequencing technologies, we aimed to analyze genomes of microorganisms from Amazonian environments in the NCBI database to characterize their BGC diversity and explore the biotechnological potential of these unique microorganisms.

MATERIAL AND METHODS

Genome selection and filtering

Genomes were retrieved from the NCBI database (https://www.ncbi.nlm.nih.gov/) in August 2023. A search was conducted in the nucleotide database using the keyword "Amazon [All_Fields]", filtered for fungal and bacterial sequences longer than 500,000 base pairs. Each genome (complete or draft) was individually evaluated, and those with metadata indicating an isolation source (or related metadata) directly associated with humans, domestic animals, or parasites of these hosts were excluded. Detailed characteristics of the selected organisms and genomes, as reported in their NCBI entries, are presented in Tables 1 and 2.

Table 1. Amazonian microorganisms with genomes available in the NCBI database. Genomes were obtained either from metagenomic assemblies (Code ending in "1") or from cultured isolates (Code ending in "2"). Data was retrieved from the respective NCBI entries and from literature references associated with the genomes, identified via PubMed searches using the species and strain identifiers.

Code	Organism	Accession code	Amazonian source	Reference
A1	*Afipia sp.* Bin.003_Pasture	GCA_024508675	Pasture soil	Mandro *et al*. 2022
B1	*Agrobacterium sp*. Bin.012_Pasture	GCA_024508975	Pasture soil	Mandro *et al*. 2022
C1	*Chitinophagaceae* Bin.005_Pasture	GCA_024508725	Pasture soil	Mandro *et al*. 2022
D1	*Flavipsychrobacter sp*. Bin.009_Pasture	GCA_024508935	Pasture soil	Mandro *et al*. 2022
E1	*Massilia sp.* Bin.003_Forest	GCA_024508955	Forest soil	Mandro *et al*. 2022
F1	*Mesorhizobium sp* Bin.005_Forest	GCA_024509055	Forest soil	Mandro *et al*. 2022
G1	*Nocardioides* Bin.006_Pasture	GCA_024509035	Pasture soil	Mandro *et al*. 2022
H1	*Nocardioides kongjuensis* Bin.001_Pasture	GCA_024508815	Pasture soil	Mandro *et al*. 2022
I1	*Nocardioides kongjuensis* Bin.002_Forest	GCA_024508895	Forest soil	Mandro *et al*. 2022
J1	*Paenarthrobacter sp*. Bin.011_Pasture	GCA_024508985	Pasture soil	Mandro *et al*. 2022
K1	*Paraburkholderia sp* Bin.013_Forest	GCF_024281035	Forest soil	Mandro *et al*. 2022
L1	Phenylobacterium sp. Bin.007_Pasture	GCA_024508875	Pasture soil	Mandro *et al*. 2022
M1	Ramlibacter sp. Bin.015_Pasture	GCA_024508995	Pasture soil	Mandro *et al*. 2022
N1	Rhodanobacteraceae Bin.008_Pasture	GCA_024508655	Pasture soil	Mandro *et al*. 2022
O1	Sphingobium sp. UBA5915	GCA_002431195	Amazon River	Parks *et al*. 2017
P1	Sphingomonas sp. Bin.002_Pasture	GCF_024508665	Pasture soil	Mandro *et al*. 2022
Q1	Sphingomonas sp. Bin.006_Forest	GCF_024508715	Forest soil	Mandro *et al*. 2022
R1	Xanthobacteraceae Bin.001_Forest	GCA_024508855	Forest soil	Mandro *et al*. 2022
S1	Xanthomonadaceae Bin.004_Pasture	GCA_024508915	Pasture soil	Mandro *et al*. 2022
A2	*Acinetobacter junii* SB132	GCA_004123275	Aquatic macrophyte	Unpublished
B2	*Aquitalea* sp. LB_tupeE	GCF_013391415	Lake	Castro *et al*. 2021
C2	*Bacillus paramycoides* LB_RP2	GCA_013391375	Lake	de Castro *et al*. 2021
D2	*Bacillus velezensis* P34	GCF_005696395	Gut of Amazon River fish	Stincone *et al*. 2020
E2	*Bradyrhizobium campsiandrae* INPA384B	GCA_014530645	Amazonian plant nodules	Cabral Michel *et al*. 2021
F2	*Bradyrhizobium campsiandrae* INPA394B	GCA_014529705	Amazonian plant nodules	Cabral Michel *et al*. 2021
G2	*Bradyrhizobium forestalis* INPA54B	GCA_002795245	Amazon forest legume species	Martins da Costa *et al*. 2018
H2	*Bradyrhizobium manausense* R3351	GCA_001440035	Amazonian plant nodules	Silva *et al*. 2014
I2	*Bradyrhizobium sp* UFLA03-84	GCF_002289535	Amazonian plant nodules	Martins da Costa *et al*. 2019
J2	*Candidatus Mycobacterium wuenschmannii*	GCA_030252325	Amazon milk frogs	Zeineldin *et al*. 2023
K2	*Mucilaginibacter_sp.* PPGCB2223	GCA_001705515	Pasture soil	Unpublished
L2	*Nostoc* sp. CMAA1605	GCA_021596505	Amazon River	Unpublished
M2	*Nostoc piscinale* CENA21	GCF_001298445	Wet sediment, Mari-Mari Island	Leão *et al*. 2016
N2	*Pseudomonas putida* IEC33019	GCA_002741075	Aurá River	Unpublished
O2	*Pseudonocardia sp.* ICBG1034	GCA_020131375	Forest-insect association	Fukuda *et al*. 2021
P2	*Pseudonocardia sp.* ICBG1293	GCA_020131355	Forest-insect association	Fukuda *et al*. 2021
Q2	*Pseudonocardia sp.* ICBG601	GCA_020496245	Forest-insect association	Fukuda *et al*. 2021
R2	*Ralstonia pickettii* LB_tupeA	GCA_013391365	Lake	Unpublished
S2	*Rickettsia amblyommatis* AC37	GCF_001273795	Forest-tick association	Yen *et al*. 2021
T2	*Sinomonas sp.* R1AF57	GCF_002224485	Forest soil	Unpublished
U2	*Synechococcus sp. GFB01*	GCA_001039265	Freshwater lagoon	Guimarães *et al*. 2015

Table 2. Standard genome quality metrics and assembly metrics of Amazonian microorganisms with genomes available in the NCBI database. Parameters were retrieved from NCBI datasets or estimated with CheckM. GS = genome size; N50 = statistical metric that indicates the assembly quality by identifying the length at which 50% of the genome assembly is contained in contigs of that length or longer.

Code	GS (MB)	Contigs	N50 (pb)	Completness (%)	Contamination (%)
A1	4.4	23	562,041	67.58	19.98
B1	5.3	1395	6,412	67.98	8.92
C1	3.9	19	1,314,444	99.42	1.18
D1	3.8	254	94,077	99.27	1.45
E1	7.5	77	228,925	94.39	14.11
F1	6.7	588	20,114	95.90	8.36
G1	4.8	453	16,445	95.57	2.24
H1	5.5	22	494,217	80.06	2.15
I1	5.5	12	531,985	80.10	2.15
J1	3.2	2527	1,758	56.13	1.20
K1	5.6	78	141,664	99.23	0.98
L1	4.8	375	21,956	94.03	4.87
M1	4.1	1470	4,269	77.44	0.14
N1	4.7	222	37,337	99.55	0.33
O1	4.6	184	51,982	91.79	8.35
P1	4.6	8	1,260,393	98.81	3.50
Q1	3.2	93	83,256	91.89	1.94
R1	5.2	113	151,135	99.83	0.55
S1	2.4	158	21,631	99.77	0.12
A2	3.5	63	138,073	97.21	0.97
B2	4.5	68	207,132	99.78	1.79
C2	5.3	76	216,793	98.82	0.22
D2	3.9	1	3,153,971	99.41	0.00
E2	8.9	172	109,464	98.82	0.45
F2	9.1	483	110,456	99.31	1.04
G2	8.3	189	151,515	99.91	0.32
H2	9.1	371	45,300	96.51	1.43
I2	1.5	6	832,094	99.96	0.45
J2	5.1	1	5,102,433	98.65	0.54
K2	5.5	8	970,906	99.13	1.82
L2	6.7	192	56,867	85.73	1.22
M2	7.1	1	1,576,034	89.25	1.30
N2	5.8	1	5,794,410	97.34	2.56
O2	6.7	32	1,138,390	92.87	7.69
P2	8.0	86	418,257	91.25	8.70
Q2	13.3	280	143,282	94.67	63.7
R2	5.3	11	1,205,870	89.48	5.99
S2	1.4	1	1,415,369	99.38	0.30
T2	3.9	1	262,651	99.32	1.31
U2	2.3	125	29,091	68.51	2.30

Identification of BGCs

To identify biosynthetic gene clusters (BGCs) related to secondary metabolites, two computational tools were used: Antibiotics & Secondary Metabolite Analysis Shell (AntiSMASH 7.0) and Prediction Informatics for Secondary Metabolomes (Prism 4.4.5) (Skinnider *et al*. 2015; Blin *et al*. 2023). AntiSMASH was executed with relaxed detection strictness, and all additional features were enabled. Prism analysis was executed with default parameters, which generated compound predictions in SMILES (Simplified Molecular-Input Line-Entry System) format. Each SMILES string was used to perform an exact structure-based search in the PubChem database (https://pubchem.ncbi.nlm.nih.gov/), accessed in October 2023, to identify the corresponding compound. The matches were classified based on the compound identities retrieved from the identity tab in PubChem.

As one of the features, the results obtained from AntiSMASH were compared with the MiBIG database (Minimum Information about a Biosynthetic Gene Cluster, https://mibig.secondarymetabolites.org/). Predicted compounds and known clusters were categorized according to their similarity. Compounds with high similarity (>1.2 for compound and >75% for gene clusters) were evaluated for potential bioactivity using available literature.

BGC similarity clustering and comparison of protein sequences

Initially, BGCs identified were labeled as PRISM standard. The AntiSMASH outputs were further analyzed using the Biosynthetic Gene Similarity Clustering And Prospecting Engine (BiG-SCAPE) version 2.0 (Navarro-Muñoz *et al*. 2020) in combination with the MiBIG version 4.0 database (Zdouc *et al*. 2025) and Pfam version 36.0 (Paysan-Lafosse *et al*. 2025) using a cutoff value of 0.3. All analyses were performed using Python 3.11 via Docker (Docker Desktop version 28.1), with key packages including Biopython 1.85 (Cock *et al*. 2009).

BGCs from Amazon microorganisms and MiBIG entries predicted to be closely related based on Pfam domains had their protein sequences extracted from antiSMASH-generated .gbk files and the MiBIG repository. Sequences were converted to FASTA format using Biopython. Protein sequences from each BGC were compared pairwise using global alignments with the Needleman-Wunsch algorithm, as implemented in the pairwise2.align.globalxx function from the BioPython library. Percent identity was calculated based on the alignment score divided by the length of the longer sequence in each pair.

RESULTS

The search in the NCBI database yielded 21 genomes of isolated microorganisms and 19 genomes of microorganisms

assembled from metagenome sequencing. Among the isolated microorganisms, six genomes were complete (i.e., fully sequenced). In total, 40 genomes were selected and analyzed, all belonging to the bacterial domain. Although bacterial and fungal genomes were initially identified, all fungal genomes, as well as bacterial genomes were excluded based on the predefined criteria. Among the 40 genomes, bacterial diversity comprised only five phyla and eight classes: 23 belonged to the phylum Pseudomonadota, nine to Actinomycetota, three to Bacteroidota, three to Cyanobacteriota, and two to Bacillota (Table 3).

AntiSMASH and PRISM predicted a total of 597 biosynthetic gene clusters (BGCs); PRISM predicted 195 and AntiSMASH 402 (Figure 1). PRISM does not group ribosomally synthesized and post-translationally modified peptides (RiPPs), so, for consistency, the 12 lassopeptides, 12 bacteriocins, and nine lantipeptides were grouped into the RiPP category (Figure 1a,b). PRISM did not predict BGCs for genomes M1 (*Ramlibacter sp.*), K2 (*Mucilaginibacter sp.*), S2 (*Rickettsia amblyommatis*), and U2 (*Synechococcus sp.),* while only genome S2 lacked predictions in AntiSMASH.

Genome quality varied across the dataset (Table 2), which likely influenced BGC predictions. Notably, some genomes with high contamination or fragmented assemblies (such as E1, L2, and Q2) had a relatively high number of predicted BGCs, which may reflect assembly artifacts. Conversely, M1 presented low completeness, high fragmentation and yielded fewer BGCs. This highlights the need to consider genome quality when interpreting biosynthetic potential from genomes.

The comparative analysis revealed that AntiSMASH predicted more BGCs than PRISM, except for homoserine lactone, polyketides, and siderophores (Figure 2). These differences likely result from variations in both the underlying algorithms and their respective databases. For instance, genomes M1, K2, and U2 had no BGCs predicted by PRISM, but yielded four, eight, and three predicted BGCs, respectively, using AntiSMASH.

Comparing the predicted BGCs with MIBiG reference clusters revealed that most BGCs predicted natively by AntiSMASH showed low similarity to MIBiG entries. Specifically, nine had similarity scores > 1.0; 49 had scores between 0.5 and 1.0; 136 between 0.3 and 0.5; and 208 had scores < 0.3. Among those with a high similarity score (>1.0)

Table 3. Taxonomic distribution of Amazonian microorganisms with genomes available in the NCBI database based on genome origin. N met = number of metagenomes; N iso = isolates.

Phylum / Class	N met	N iso
Pseudomonadota / Alphaproteobacteria	8	6
Pseudomonadota / Betaproteobacteria	3	2
Pseudomonadota / Gammaproteobacteria	2	2
Actinomycetota / Actinomycetes	4	5
Bacteroidota / Chitinophagia	2	0
Bacteroidota / Sphingobacteriia	0	1
Bacillota / Bacilli	0	2
Cyanobacteriota / Cyanophyceae	0	3
Total	**19**	**21**

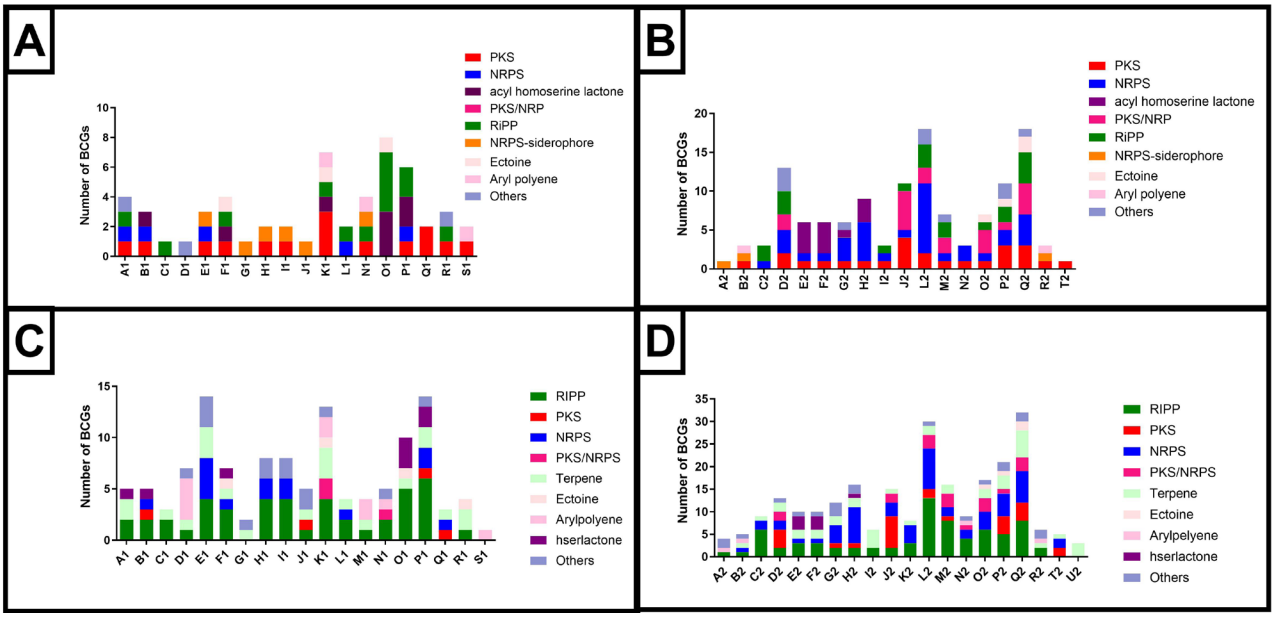


Figure 1. Biosynthetic gene clusters (BGCs) of secondary metabolites predicted from Amazonian microbial genomes using PRISM (A and C) and AntiSMASH (B and D). (A) and (C) show genomes assembled from metagenome data, while (B) and (D) represent genomes from isolated microorganisms. A1-S1 and A2-U2 correspond to the microorganisms listed in Table 1.

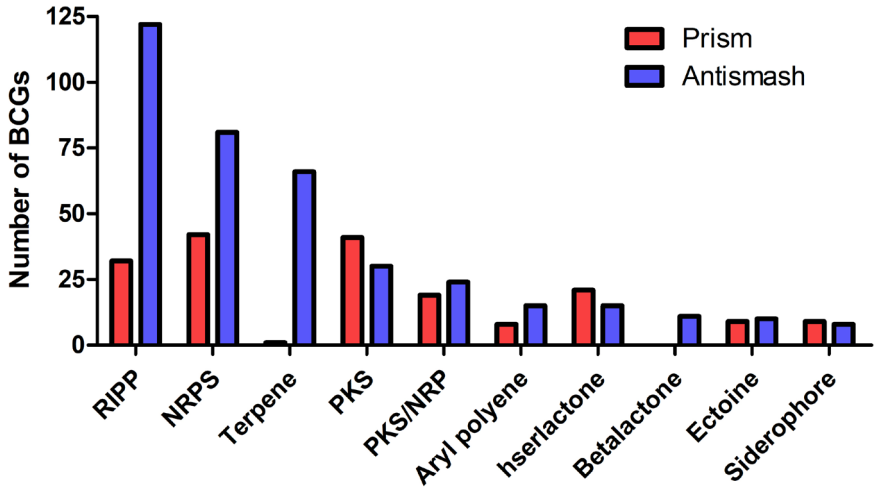


Figure 2. Number of predominant biosynthetic gene cluster (BGC) classes predicted from Amazonian microbial genomes by PRISM and AntiSMASH used in the present study.

(Table S1), three predicted compounds from the *Bacillus velezensis* (D2) genome are especially relevant for their high similarity to known antimicrobials: one BGC showed 1.67 similarity to the non-ribosomal peptide/polyketide synthase (NRP/PKS) bacillomycin (Fazle Rabbee and Baek 2020); another predicted pathway showed a similarity of 1.68 with the NRP/PKS bacillaene (Fazle Rabbee and Baek 2020); and a third BGC exhibited a similarity of 2.61 with the RiPP antibiotic amylocyclicin (Scholz *et al*. 2014). In addition to the three BGCs predicted from *Bacillus velezensis*, the *Nostoc sp*. CMAA1605 genome contained a BGC with a similarity score of 1.44 to shinorine, a compound with UV-protective properties used in biodegradable sunscreens (Jin *et al*. 2021). Another compound with a 1.24 similarity to the immunosuppressant thalassospiramide (Oh *et al*. 2007) was predicted from the *Pseudonocardia sp*. ICBG1034 genome. BGCs showing high similarity (>75%) were also detected with antifungal (Ye *et al*. 2024), antimicrobial (Kanda *et al*. 1975; Ye *et al*. 2013; Dose *et al*. 2018; Grigoreva *et al*. 2021), antitumoral (Salis *et al*. 2014), protein inhibition (Rohrlack *et al*. 2004; Harms *et al*. 2016; Jokela *et al*. 2017), chelating (Harris *et al*. 2017; Shankar and Akhter 2024), and other activities (Hoffmann *et al*. 2003; Tobias *et al*. 2018; Arul Prakash and Kamlekar 2021; Parihar *et al*. 2022; Sajnaga *et al*. 2024) (Table 4). We also identified clusters with potential to produce carotenoids and ectoine, compounds with biotechnological relevance in food, agriculture, cosmetics, and medicine (Table 4) (López *et al*. 2023; Šišić *et al*. 2024).

The Prism/PubChem analysis did not return any exact compound matches using the SMILES prediction. As output files generated by PRISM are incompatible with BiG-SCAPE, the clustering analysis was based exclusively on antiSMASH predictions. Because BiG-SCAPE is more selective and focuses on well-characterized BGCs suitable for clustering and analysis, it yielded fewer BGCs (366) than AntiSMASH (402). BIG-SCAPE categorized the 366 BGCs by biosynthetic category (Figure 3) and cluster class (Table S2-S9). The most frequently observed categories were RiPP, NRPS, and terpene (Table S10). In class-based analysis, terpenes, RiPP-like, NRPS, and homoserine lactones were predominant. About 75% of the 366 BGCs were classified as singletons (unique clusters), while only 12 matched known entries in the MIBiG database (Table 5). Network layouts generated by BiG-SCAPE for each BGC class (Figure S1) illustrate the overall diversity and novelty of the biosynthetic potential, with numerous singletons and novel cluster families identified across the analyzed genomes.

Of the predicted chemical structures of the12 BGCs with known MIBiG clusters (Figure 4), eight were already described (Table 4, marked with *). Two clusters (BGC0000869 heterocyst glicolipids and BGC0001748 Pseudospumigin) had no available structures in the MiBIG repository. Pairwise global alignments between the BGCs and their corresponding MIBiG reference sequences showed overall high similarity, with many regions exceeding 97%, and conserved gene organization, supporting the reliability of BGC identification (Figures S2–S13).

DISCUSSION

We unexpectedly identified several genomes from the Amazon of metagenomic origin in NCBI that were sequenced from mixed DNA extracted from microbial communities, posing challenges for accurate genome assembly and completeness. Furthermore, metagenome-derived genomes showed limited representativeness, as 18 out of 19 originated from a single

Table 4. Biosynthetic gene clusters of Amazonian microorganisms with genomes available in the NCBI database with high similarity (>75%) to known clusters in MiBIG database. NRP = nonribosomal peptides; PKS = polyketide; RiPP = ribosomally synthesized and post-translationally modified peptides.

Code	Most similar known cluster	Type	Activity	Reference
A1	N-acyl alanine	Other	multiples	Arul Prakash and Kamlekar 2021
B2	desferrioxamine E	Other	Siderophore / Antitumor	Salis *et al*. 2014
C2	Bacillibactin*	NRP	Antimicrobial	Fazle Rabbee and Baek 2020
D2	Andalusicin	RiPP	Antibiotic	Grigoreva *et al*. 2021
D2	Surfactin*	NRP	Antimicrobial	Fazle Rabbee and Baek 2020
D2	macrolactin H*	PKS	Antimicrobial	Fazle Rabbee and Baek 2020
D2	Bacillaene*	PKS	Antimicrobial	Fazle Rabbee and Baek 2020
D2	Fengycin*	NRP	Antifungal	Fazle Rabbee and Baek 2020
D2	Difficidin*	PKS	Antimicrobial	Fazle Rabbee and Baek 2020
D2	Bacillibactin*	NRP	Antimicrobial	Fazle Rabbee and Baek 2020
D2	Bacilysin*	Other	Antimicrobial	Fazle Rabbee and Baek 2020
E1	rhizomide	NRP	Antifungal	Ye *et al*. 2024
E1	carotenoid	Terpene	Multiples	López *et al*. 2023
G2	gamexpeptide C	NRP	Insecticide	Tobias *et al*. 2018
G2	rhizomide	NRP	Antifungal	Ye *et al*. 2024
H1	ε-Poly-L-lysine	NRP	Antimicrobial	Ye *et al*. 2013
H1	ε-Poly-L-lysine	NRP	Antimicrobial	Ye *et al*. 2013
H2	kolossin	NRP	Antiprotozoa	Parihar *et al*. 2022
H2	icosalide	NRP	Antibiotic	Dose *et al*. 2018
H2	ririwpeptide	NRP	Unknown	Sajnaga *et al*. 2024
J2	mycobactin	NRP.PKS	siderophore	Shankar and Akhter 2024
J2	isonitrile lipopeptides	NRP	Metallophores	Harris *et al*. 2017
J2	alkylresorcinol	PKS	Antibiotic	Kanda *et al*. 1975
L2	shinorine/ 4-deoxygadusol/ mycosporine glycine	NRP	Sunscreen	Jin *et al*. 2021
L2	Pseudospumigin*	NRP.PKS	Protease inhibitor	Jokela *et al*. 2017
L2	anabaenopeptin	NRP	Protease inhibitor	Harms *et al*. 2016
M2	microviridin	RiPP	Protease inhibitor	Rohrlack *et al*. 2004
M2	nostopeptolide	PKS.NRP	Unknown	Hoffmann *et al*. 2003
O2	ectoine	Other	Multiples	Šišić *et al*. 2024
P2	ectoine	Other	Multiples	Šišić *et al*. 2024
Q2	ectoine	Other	Multiples	Šišić *et al*. 2024
R1	ectoine	Other	Multiples	Šišić *et al*. 2024

* Confirmed in BIG-SCAPE analysis against MIBiG database

Table 5. Summary of BiG-SCAPE clustering results for biosynthetic gene clusters from Amazonian microorganisms with genomes available in the NCBI database. Each family with a MiBIG match is illustrated in Figures S2–S13.

Category	BGCs	Singletons	Protein families	Families with MIBiG
NRPS	61	46	52	2
NRPS.other	3	3	3	0
NRPS.PKS	14	9	11	2
NRPS.PKS.other	1	1	1	2
NRPS.RiPP	7	7	7	2
NRPS.terpene	1	1	1	0
other	74	52	61	2
PKS	27	27	27	3
PKS.other	1	1	1	0
PKS.RiPP	1	1	1	0
RiPP	114	85	97	0
RiPP.other	2	2	2	0
saccharide	1	0	1	0
terpene	59	39	47	0
Total	366	276	312	12

research project conducted in forest soil or pasture (Mandro *et al*. 2022), with only one genome assembled from a river-sample metagenome (O1) (Parks *et al*. 2017). While the genomes derived from isolated microorganisms represented more diverse organisms or environments, including insects (Fukuda *et al*. 2021; Yen *et al*. 2021), river-associated sediment (Leão *et al*. 2016), fishes (Stincone *et al*. 2020), frogs (Zeineldin *et al*. 2023), plants (Silva *et al*. 2014; Martins da Costa *et al*. 2018; Cabral Michel *et al*. 2021), lakes (Guimarães *et al*. 2015; Castro *et al*. 2021. de Castro *et al*. 2021). These genomes represent only a fraction of the immense microbial diversity of the Amazon (Buscardo *et al*. 2018; Thompson *et al*. 2023). Although genome completeness, contamination, N50, and contig number may limit the recovery of complete BGCs, their identification can still offer valuable insights for bioprospecting and the exploration of microbial metabolic potential.

The prediction of secondary metabolic biosynthesis pathways is particularly significant due to the potential

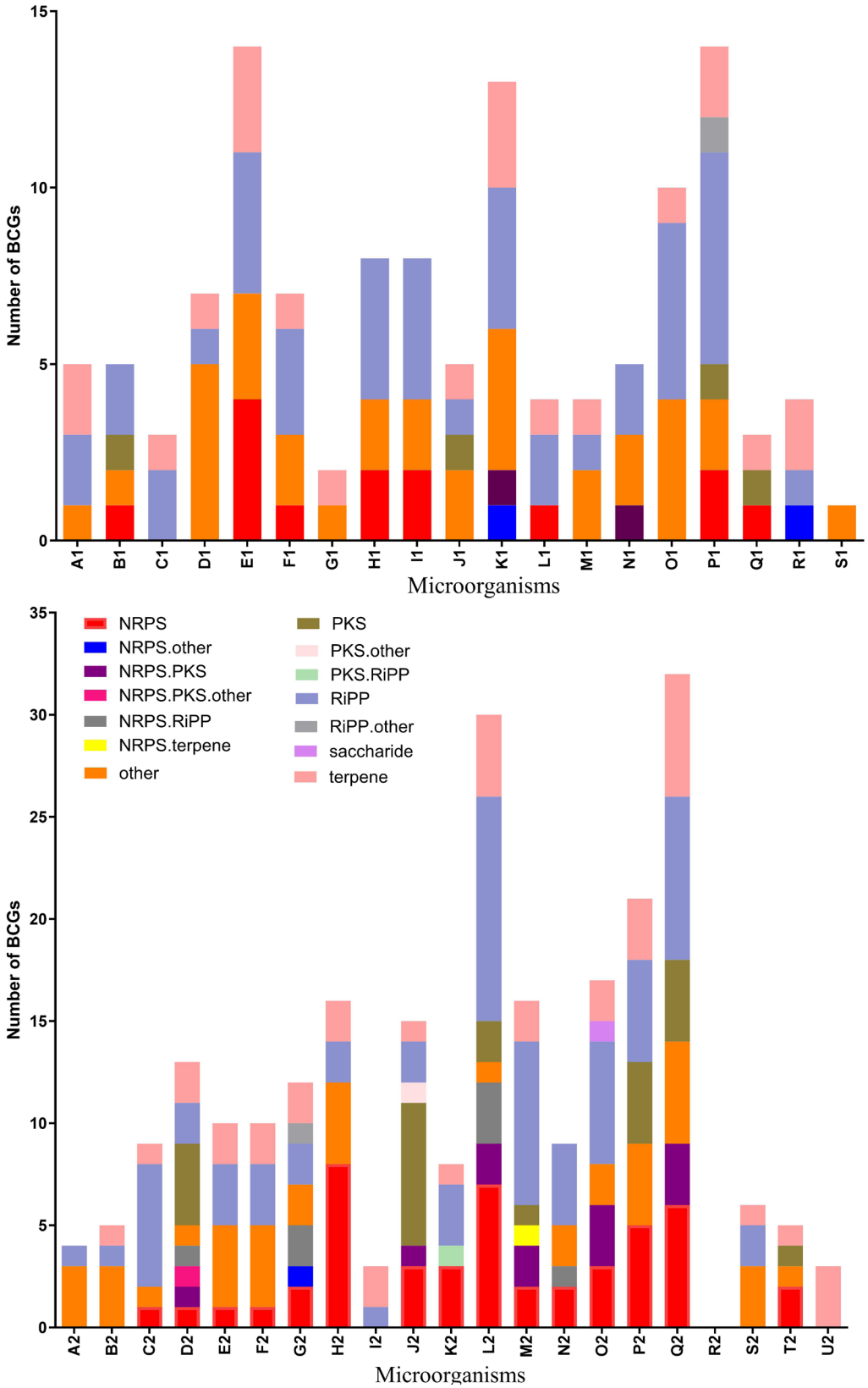


Figure 3. Classification and distribution of biosynthetic gene clusters (BGCs) predicted from Amazonian microbial genomes using BIG-SCAPE. A1-S1 and A2-U2 correspond to the microorganisms listed in Table 1.

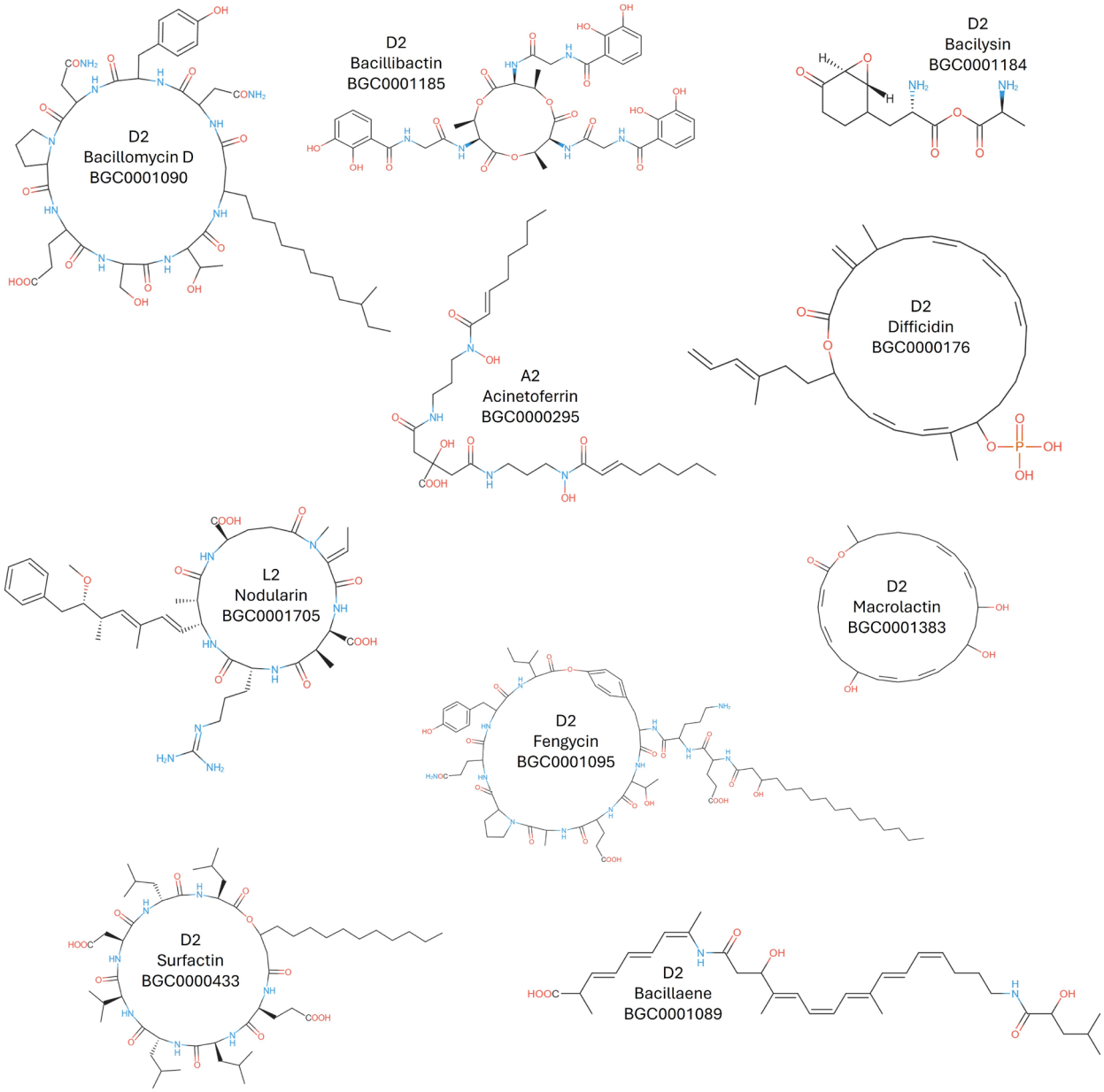


Figure 4. Chemical structure of the closest secondary metabolites associated with biosynthetic gene clusters (BGCs) from Amazonian microorganisms treated in this study, based on BiG-SCAPE similarity to characterized BGCs in the MIBiG repository.

for exploring these microorganisms for biotechnological purposes. The high similarities observed in AntiSMASH/MiBIG of the aforementioned BGCs highlight their capacity to synthesize valuable bioactive compounds that are either identical to or chemically similar to established molecules, as exemplified by bacillomycin (Zhou *et al.* 2018). These clusters may thus be linked to known classes or analogs with proven applications. The analogy of the most similar compounds (with scores from 0.5 to 1.0) to known secondary metabolites with antifungal activity, such as amphotericin B and siderophores like mycobactin, highlights the unexplored potential of Amazonian microorganisms for secondary metabolite production. However, genome mining based on database-derived samples has inherent limitations, as it does not account for the actual expression or production of these compounds under *in vivo* conditions.

Although the biotechnological potential of Amazonian microorganisms has been addressed in previous studies, most of this knowledge derives from relatively few isolates and from studies targeting cultivable taxa (Cabral Michel *et al.* 2021; Cardona *et al.* 2022; Lopes *et al.* 2025). In contrast,

our genome mining approach includes both cultured isolates and metagenome-assembled genomes, revealing biosynthetic potential in a broader range of microbial lineages.

One advantage of genome mining relies on its applicability to any sequenced genomes, however, the identification of BGCs is often limited to those that are already known (Wohlleben *et al*. 2016). This limitation implies that additional BGCs, particularly from less commonly studied microorganisms, may go undetected, leaving their associated secondary metabolites undiscovered. Further studies should involve inducing the production of secondary metabolites by microbial isolates (the sources of the genomes) combined with information from analysis such as AntiSMASH, which can also predict potential regulators of BGC expression (Bhattacharjee *et al*. 2023; Blin *et al*. 2023). Novel technologies, including synthetic biology and genome editing, offer the potential to evaluate the production of these BGCs in other organisms by synthesizing and cloning, and to stimulate secondary metabolite production in a host, thus assessing their activity *in vivo* (Choi *et al*. 2018), or even by using synthetic biological cells with the BGC introduced into the genome (Kunakom and Eustáquio 2019).

Among the 21 isolates analyzed, only *Pseudonocardia sp*. was previously shown to produce an antifungal compound (attinimicin) *in vivo,* for which a BGC was predicted from its genome (Fukuda *et al*. 2021). Although the attinimicin BGC is not listed in the MiBIG database, our analysis revealed the gene cluster organization associated with attinimicin in our AntiSMASH data and BIG-SCAPE cluster output (Figure S14). This case reinforces the potential of low-similarity BGCs, especially NRP, RiPP, and polyketides (Wenski *et al*. 2022) that were commonly found in those genomes. Notably, *Pseudonocardia* isolates were obtained associated with ants in the forest. Thus, the rich diversity of Amazonian plants and animals may harbor equally diverse microbial communities and new BCGs and their metabolites not only in soil and water, but also associated with the biota (Pereira *et al*. 2017).

The high number of singletons and unique protein families identified by BiG-SCAPE further emphasizes the biosynthetic novelty harbored within Amazonian microbes. Even with the relatively small dataset of only 40 genomes, we uncovered substantial novelty in gene cluster composition and diversity. This finding highlights the urgent need for expanded genomic surveys of microbial communities across Amazonian ecosystems.

Finally, the predicted BGCs hold promise in addressing global challenges such as antibiotic resistance and environmental sustainability. Several clusters showed similarity to known antibacterial compounds, such as macrolactins and bacillaenes, which are associated with inhibition of Gram-positive pathogens and microbial competition (Fazle Rabbee and Baek 2020). This suggests a potential source of novel antibiotics that could help combat multidrug-resistant bacteria. Additionally, we identified clusters related to biosurfactants and siderophores, which could be utilized as antimicrobials and also in environmentally friendly biocontrol strategies and bioremediation. Examples include the biosurfactant surfactin and the siderophore bacillibactin (Fazle Rabbee and Baek 2020). These compounds offer opportunities for sustainable applications, including biodegradable alternatives to chemical pesticides or pollutants in agriculture and bioremediation, an important approach to minimizing environmental impact in ecologically sensitive and biologically rich regions such as the Amazon.

CONCLUSIONS

This study highlights the utility and relevance of *in silico* genome analysis for the identification of BGCs associated with secondary metabolite production in microorganisms from the highly biodiverse Amazon region. The analysis of 40 genomes, covering various bacterial phyla, revealed considerable biosynthetic potential, including the detection of BGCs similar to known bioactive compounds, in addition to many clusters with low or no similarity to reference databases, suggesting the presence of novel metabolic pathways. However, our BGC predictions are limited by the available reference data, especially for underexplored taxa typical of the Amazonian microbiota. Furthermore, the functional characterization and validation of these metabolites depend on experimental *in vivo* studies, which remain scarce for many of the species analyzed. This work highlights the value of genome mining in revealing the hidden metabolic potential of Amazonian microorganisms and reinforces the need for expanded genomic and experimental efforts to fully harness their biotechnological potential. Current evidence, including our findings, suggests that these microorganisms represent an underexplored and promising source of bioactive compounds.

ACKNOWLEDGMENTS

The study was conducted without third-party funding. ELSM held a technological development fellowship (DTI) from Fundação de Amparo à Pesquisa do Amazonas (FAPEAM) during the revision period.

REFERENCES

Almeida Lima, B.; Dos Santos Scherer, R.; Brigatto Albino, U.; Da Silva Siqueira, F.F.; Pires Bitencourt, J.A.; Cerqueira dos Santos, S. 2023. Atividade enzimática e antimicrobiana de microrganismos isolados de uma caverna da região Amazônica. *Scientia Plena* 19:1-14. doi: 10.14808/sci.plena.2023.046201.

Arul Prakash, S.; Kamlekar, R.K. 2021. Function and therapeutic potential of N-acyl amino acids. *Chemistry and Physics of Lipids* 239: 105114.

Bauman, K.D.; Butler, K.S.; Moore, B.S.; Chekan, J.R. 2021. Genome mining methods to discover bioactive natural products. *Natural Product Reports* 38: 2100–2129.

Bhattacharjee, A.; Sarma, S.; Sen, T.; Devi, M.V.; Deka, B.; Singh, A.K. 2023. Genome mining to identify valuable secondary metabolites and their regulation in Actinobacteria from different niches. *Archives of Microbiology* 205: 127. doi: 10.1007/s00203-023-03482-3

Blin, K.; Shaw, S.; Augustijn, H.E.; Reitz, Z.L.; Biermann, F.; Alanjary, M.; *et al*. 2023. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Research* 51: W46–W50.

Buscardo, E.; Geml, J.; Schmidt, S.K.; Freitas, H.; da Cunha, H.B.; Nagy, L. 2018. Spatio-temporal dynamics of soil bacterial communities as a function of Amazon forest phenology. *Scientific Reports* 8: 4382. doi: 10.1038/s41598-018-22380-z

Cabral Michel, D.; Martins da Costa, E.; Azarias Guimarães, A.; Soares de Carvalho, T.; Santos de Castro Caputo, P.; Willems, A.; *et al*. 2021. *Bradyrhizobium campsiandrae* sp. nov., a nitrogen-fixing bacterial strain isolated from a native leguminous tree from the Amazon adapted to flooded conditions. *Archives of microbiology* 203: 233–240.

Cardona, G.I.; Escobar, M.C.; Acosta-González, A.; Marín, P.; Marqués, S. 2022. Highly mercury-resistant strains from different Colombian Amazon ecosystems affected by artisanal gold mining activities. *Applied Microbiology and Biotechnology* 106: 2775–2793.

Carvalho, J.M.; Paixão, L.K.O. da; Dolabela, M.F.; Marinho, P.S.B.; Marinho, A.M. do R. 2016. Phytosterols isolated from endophytic fungus *Colletotrichum gloeosporioides* (Melanconiaceae). *Acta Amazonica* 46: 69–72.

Castro, L.M.; Foong, C.P.; Higuchi-Takeuchi, M.; Morisaki, K.; Lopes, E.F.; Numata, K.; *et al*. 2021. Microbial prospection of an Amazonian blackwater lake and whole-genome sequencing of bacteria capable of polyhydroxyalkanoate synthesis. *Polymer Journal* 53: 191–202.

de Castro, L.M.; Foong, C.P.; Higuchi-Takeuchi, M.; Lopes, E.F.; Numata, K.; Dias da Silva, S.; *et al*. 2021. Draft whole-genome sequence of *Bacillus paramycoides* LB_RP2, a putative polyhydroxyalkanoate-producing bacterium isolated from an Amazonian blackwater river. *Microbiology Resource Announcements* 10: e00438-21.

Cerqueira dos Santos, S.; Araújo Torquato, C.; de Alexandria Santos, D.; Orsato, A.; Leite, K.; Serpeloni, J.M.; *et al*. 2024. Production and characterization of rhamnolipids by *Pseudomonas aeruginosa* isolated in the Amazon region, and potential antiviral, antitumor, and antimicrobial activity. *Scientific Reports* 14: 4629. doi:10.1038/s41598-024-54828-w

Choi, S.-S.; Katsuyama, Y.; Bai, L.; Deng, Z.; Ohnishi, Y.; Kim, E.-S. 2018. Genome engineering for microbial natural product discovery. *Current Opinion in Microbiology* 45: 53–60.

Cock, P.J.A.; Antao, T.; Chang, J.T.; Chapman, B.A.; Cox, C.J.; Dalke, A.; *et al*. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25: 1422–1423.

Corral-García, L.S.; Molina, M.C.; Bautista, L.F.; Simarro, R.; Espinosa, C.I.; Gorines-Cordero, G.; *et al*. 2024. Bacterial diversity in old hydrocarbon polluted sediments of Ecuadorian Amazon river basins. *Toxics* 12: 119. doi: 10.3390/toxics12020119

da Cruz, D.C.; Benayas, J.M.R.; Ferreira, G.C.; Santos, S.R.; Schwartz, G. 2021. An overview of forest loss and restoration in the Brazilian Amazon. *New Forests* 52: 1–16.

Dose, B.; Niehs, S.P.; Scherlach, K.; Flórez, L. V.; Kaltenpoth, M.; Hertweck, C. 2018. Unexpected bacterial origin of the antibiotic icosalide: Two-tailed depsipeptide assembly in multifarious *Burkholderia* symbionts. *ACS Chemical Biology* 13: 2414–2420.

Duban, M.; Cociancich, S.; Leclère, V. 2022. Nonribosomal peptide synthesis definitely working out of the rules. *Microorganisms* 10: 577. doi: 10.3390/microorganisms10030577

Ellwanger, J.H.; Kulmann-Leal, B.; Kaminski, V.L.; Valverde-Villegas, J.M.; Veiga, A.B.G. da; Spilki, F.R.; *et al*. 2020. Beyond diversity loss and climate change: Impacts of Amazon deforestation on infectious diseases and public health. *Anais da Academia Brasileira de Ciências* 92: e20191375.

Fazle Rabbee, M.; Baek, K.-H. 2020. Antimicrobial activities of lipopeptides and polyketides of *Bacillus velezensis* for agricultural applications. *Molecules* 25: 4973. doi: 10.3390/molecules25214973

Fonseca, J.P.; Hoffmann, L.; Cabral, B.C.A.; Dias, V.H.G.; Miranda, M.R.; de Azevedo Martins, A.C.; *et al*. 2018. Contrasting the microbiomes from forest rhizosphere and deeper bulk soil from an Amazon rainforest reserve. *Gene* 642: 389–397.

Fukuda, T.T.H.; Helfrich, E.J.N.; Mevers, E.; Melo, W.G.P.; Van Arnam, E.B.; Andes, D.R.; *et al*. 2021. Specialized metabolites reveal evolutionary history and geographic dispersion of a multilateral symbiosis. *ACS Central Science* 7: 292–299.

Grigoreva, A.; Andreeva, J.; Bikmetov, D.; Rusanova, A.; Serebryakova, M.; Garcia, A.H.; *et al*. 2021. Identification and characterization of andalusicin: N-terminally dimethylated class III lantibiotic from *Bacillus thuringiensis sv. andalousiensis*. *IScience* 24: 102480.

Guimarães, P.I.; Leão, T.F.; de Melo, A.G.C.; Ramos, R.T.J.; Silva, A.; Fiore, M.F.; *et al*. 2015. Draft genome sequence of the picocyanobacterium *Synechococcus* sp. strain GFB01, isolated from a freshwater lagoon in the Brazilian Amazon. *Genome Announcements* 3: e00876-15.

Harms, H.; Kurita, K.L.; Pan, L.; Wahome, P.G.; He, H.; Kinghorn, A.D.; *et al*. 2016. Discovery of anabaenopeptin 679 from freshwater algal bloom material: Insights into the structure–activity relationship of anabaenopeptin protease inhibitors. *Bioorganic & Medicinal Chemistry Letters* 26: 4960–4965.

Harris, N.C.; Sato, M.; Herman, N.A.; Twigg, F.; Cai, W.; Liu, J.; *et al*. 2017. Biosynthesis of isonitrile lipopeptides by conserved nonribosomal peptide synthetase gene clusters in Actinobacteria. *Proceedings of the National Academy of Sciences* 114: 7025–7030.

Hetrick, K.J.; van der Donk, W.A. 2017. Ribosomally synthesized and post-translationally modified peptide natural product discovery in the genomic era. *Current Opinion in Chemical Biology* 38: 36–44.

Hoffmann, D.; Hevel, J.M.; Moore, R.E.; Moore, B.S. 2003. Sequence analysis and biochemical characterization of the nostopeptolide A biosynthetic gene cluster from *Nostoc* sp. GSV224. *Gene* 311: 171–180.

Jin, C.; Kim, S.; Moon, S.; Jin, H.; Hahn, J.-S. 2021. Efficient production of shinorine, a natural sunscreen material, from glucose and xylose by deleting HXK2 encoding hexokinase in *Saccharomyces cerevisiae*. *FEMS Yeast Research* 21: foab053.

Jokela, J.; Heinilä, L.M.P.; Shishido, T.K.; Wahlsten, M.; Fewer, D.P.; Fiore, M.F.; *et al*. 2017. Production of high amounts of hepatotoxin nodularin and new protease inhibitors pseudospumigins by the Brazilian benthic *Nostoc* sp. CENA543. *Frontiers in Microbiology* 8: 1963. doi: 10.3389/fmicb.2017.01963

Kanda, N.; Ishizaki, N.; Inoue, N.; Oshima, M.; Handa, A.; Kitahara, T. 1975. DB-2073, a new alkylresorcinol antibiotic. I. Taxonomy, isolation and characterization. *The Journal of Antibiotics* 28: 935–942.

Kormanec, J.; Novakova, R.; Csolleiova, D.; Feckova, L.; Rezuchova, B.; Sevcikova, B.; *et al*. 2020. The antitumor antibiotic mithramycin: new advanced approaches in modification and production. *Applied Microbiology and Biotechnology* 104: 7701–7721.

Kroeger, M.E.; Delmont, T.O.; Eren, A.M.; Meyer, K.M.; Guo, J.; Khan, K.; *et al*. 2018. New biological insights into how deforestation in Amazonia affects soil microbial communities using metagenomics and metagenome-assembled genomes. *Frontiers in Microbiology* 9: 1635. doi: 10.3389/fmicb.2018.01635

Kunakom, S.; Eustáquio, A.S. 2019. Natural products and synthetic biology: Where we are and where we need to go. *MSystems* 4: e00113-19.

Leão, T.; Guimarães, P.I.; de Melo, A.G.C.; Ramos, R.T.J.; Leão, P.N.; Silva, A.; *et al*. 2016. Draft genome sequence of the N2-fixing cyanobacterium *Nostoc piscinale* CENA21, isolated from the Brazilian Amazon floodplain. *Genome Announcements* 4: e00189-16.

Levine, N.M.; Zhang, K.; Longo, M.; Baccini, A.; Phillips, O.L.; Lewis, S.L.; *et al*. 2016. Ecosystem heterogeneity determines the ecological resilience of the Amazon to climate change. *Proceedings of the National Academy of Sciences* 113: 793–797.

Li, S.; Yang, B.; Tan, G.-Y.; Ouyang, L.-M.; Qiu, S.; Wang, W.; *et al*. 2021. Polyketide pesticides from actinomycetes. *Current Opinion in Biotechnology* 69: 299–307.

Lima, R.B.S.; Rocha e Silva, L.F.; Melo, M.R.S.; Costa, J.S.; Picanço, N.S.; Lima, E.S.; *et al*. 2015. In vitro and in vivo anti-malarial activity of plants from the Brazilian Amazon. *Malaria Journal* 14: 508. doi: 10.1186/s12936-015-0999-2

Lopes, M.J.D.S.; Cardoso, A.F.; Dias-Filho, M.B.; Gurgel, E.S.C.; da Silva, G.B. 2025. Brazilian Amazonian microorganisms: A sustainable alternative for plant development. *AIMS Microbiology* 11: 150–166.

López, G.-D.; Álvarez-Rivera, G.; Carazzone, C.; Ibáñez, E.; Leidy, C.; Cifuentes, A. 2023. Bacterial carotenoids: Extraction, characterization, and applications. *Critical Reviews in Analytical Chemistry* 53: 1239–1262.

Mandro, J.A.; Nakamura, F.M.; Gontijo, J.B.; Tsai, S.M.; Venturini, A.M. 2022. Metagenome-assembled genomes from Amazonian soil microbial consortia. *Microbiology Resource Announcements* 11: e00804-22.

Martins da Costa, E.; Azarias Guimarães, A.; Soares de Carvalho, T.; Louzada Rodrigues, T.; de Almeida Ribeiro, P.R.; Lebbe, L.; *et al*. 2018. *Bradyrhizobium forestalis* sp. nov., an efficient nitrogen-fixing bacterium isolated from nodules of forest legume species in the Amazon. *Archives of Microbiology* 200: 743–752.

Martins da Costa, E.; Soares de Carvalho, T.; Azarias Guimarães, A.; Ribas Leão, A.C.; Magalhães Cruz, L.; de Baura, V.A.; *et al*. 2019. Classification of the inoculant strain of cowpea UFLA03-84 and of other strains from soils of the Amazon region as *Bradyrhizobium viridifuturi* (symbiovar tropici). *Brazilian Journal of Microbiology* 50: 335–345.

McErlean, M.; Overbay, J.; Van Lanen, S. 2019. Refining and expanding nonribosomal peptide synthetase function and mechanism. *Journal of Industrial Microbiology and Biotechnology* 46: 493–513.

Mendes, M.M.G. da S.; Pereira, S.A.; Oliveira, R.L.; da Silva, L.A. de O.; Duvoisin Jr, S.; Albuquerque, P.M. 2015. Screening of Amazon fungi for the production of hydrolytic enzymes. *African Journal of Microbiology Research* 9: 741–748.

Navarro-Muñoz, J.C.; Selem-Mojica, N.; Mullowney, M.W.; Kautsar, S.A.; Tryon, J.H.; Parkinson, E.I.; *et al*. 2020. A computational framework to explore large-scale biosynthetic diversity. *Nature Chemical Biology* 16: 60–68.

Oh, D.-C.; Strangman, W.K.; Kauffman, C.A.; Jensen, P.R.; Fenical, W. 2007. Thalassospiramides A and B, immunosuppressive peptides from the marine bacterium *Thalassospira* sp. *Organic Letters* 9: 1525–1528.

Parihar, R.D.; Dhiman, U.; Bhushan, A.; Gupta, P.K.; Gupta, P. 2022. *Heterorhabditis* and *Photorhabdus* symbiosis: A natural mine of bioactive compounds. *Frontiers in Microbiology* 13: 790339.

Parks, D.H.; Rinke, C.; Chuvochina, M.; Chaumeil, P.-A.; Woodcroft, B.J.; Evans, P.N.; *et al*. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology* 2: 1533–1542.

Paysan-Lafosse, T.; Andreeva, A.; Blum, M.; Chuguransky, S.R.; Grego, T.; Pinto, B.L.; *et al*. 2025. The Pfam protein families database: embracing AI/ML. *Nucleic Acids Research* 53: D523–D534.

Pereira, J.O.; de Souza, A.Q.L.; de Souza, A.D.L.; de Castro França, S.; de Oliveira, L.A. 2017. Overview on biodiversity, chemistry, and biotechnological potential of microorganisms from the Brazilian Amazon. In: de Azevedo, J.; Quecine, M. (Eds.). *Diversity and Benefits of Microorganisms from the Tropics*, Springer International Publishing, Cham, p.71–103.

Rodrigues, R. de S.; Souza, A.Q.L. de; Barbosa, A.N.; Santiago, S.R.S. da S.; Vasconcelos, A. dos S.; Barbosa, R.D.; *et al*. 2024. Biodiversity and antifungal activities of Amazonian Actinomycetes isolated from rhizospheres of *Inga edulis* plants. *Frontiers in Bioscience-Elite* 16: 39. doi: 10.31083/j.fbe1604039

Rohrlack, T.; Christoffersen, K.; Kaebernick, M.; Neilan, B.A. 2004. Cyanobacterial protease inhibitor microviridin J causes a lethal molting disruption in *Daphnia pulicaria*. *Applied and Environmental Microbiology* 70: 5047–5050.

Sajnaga, E.; Kazimierczak, W.; Karaś, M.A.; Jach, M.E. 2024. Exploring *Xenorhabdus* and *Photorhabdus* nematode symbionts in search of novel therapeutics. *Molecules* 29: 5151. doi: 10.3390/molecules29215151

Salis, O.; Bedir, A.; Kilinc, V.; Alacam, H.; Gulten, S.; Okuyucu, A. 2014. The anticancer effects of desferrioxamine on human breast adenocarcinoma and hepatocellular carcinoma cells. *Cancer Biomarkers* 14: 419–426.

Santamaria, G.; Liao, C.; Lindberg, C.; Chen, Y.; Wang, Z.; Rhee, K.; *et al*. 2022. Evolution and regulation of microbial secondary metabolism. *ELife* 11: e76119.

Santos-Júnior, C.D.; Sarmento, H.; de Miranda, F.P.; Henrique-Silva, F.; Logares, R. 2020. Uncovering the genomic potential of the Amazon River microbiome to degrade rainforest organic matter. *Microbiome* 8: 151. doi: 10.1186/s40168-020-00930-w

Scholz, R.; Vater, J.; Budiharjo, A.; Wang, Z.; He, Y.; Dietel, K.; *et al*. 2014. Amylocyclicin, a novel circular bacteriocin produced by *Bacillus amyloliquefaciens* FZB42. *Journal of Bacteriology* 196: 1842–1852.

Shankar, G.; Akhter, Y. 2024. Stealing survival: Iron acquisition strategies of *Mycobacterium tuberculosis*. *Biochimie* 227: 37–60.

Sharrar, A.M.; Crits-Christoph, A.; Méheust, R.; Diamond, S.; Starr, E.P.; Banfield, J.F. 2020. Bacterial secondary metabolite biosynthetic potential in soil varies with phylum, depth, and vegetation type. *MBio* 11: e00416-20.

Silva, F. V; De Meyer, S.E.; Simões-Araújo, J.L.; Barbé, T. da C.; Xavier, G.R.; O'Hara, G.; *et al*. 2014. *Bradyrhizobium manausense* sp. nov., isolated from effective nodules of *Vigna unguiculata* grown in Brazilian Amazonian rainforest soils. *International Journal of Systematic and Evolutionary Microbiology* 64: 2358–2363.

Šišić, M.; Jurin, M.; Šimatović, A.; Vujaklija, D.; Jakas, A.; Roje, M. 2024. Application of biotechnology and chiral technology methods in the production of ectoine enantiomers. *Applied Sciences* 14: 8353. doi: 10.3390/app14188353

Skinnider, M.A.; Dejong, C.A.; Rees, P.N.; Johnston, C.W.; Li, H.; Webster, A.L.H.; *et al*. 2015. Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM). *Nucleic Acids Research* 43: 9645–9662.

Stincone, P.; Veras, F.F.; Pereira, J.Q.; Mayer, F.Q.; Varela, A.P.M.; Brandelli, A. 2020. Diversity of cyclic antimicrobial lipopeptides from *Bacillus* P34 revealed by functional annotation and comparative genome analysis. *Microbiological Research* 238: 126515. doi: 10.1016/j.micres.2020.126515

Thompson, C.C.; Tschoeke, D.; Coutinho, F.H.; Leomil, L.; Garcia, G.D.; Otsuki, K.; *et al*. 2023. Diversity of microbiomes across a 13,000-year-old Amazon sediment. *Microbial Ecology* 86: 2202–2209.

Tobias, N.J.; Shi, Y.-M.; Bode, H.B. 2018. Refining the natural product repertoire in entomopathogenic bacteria. *Trends in Microbiology* 26: 833–840.

Wei, B.; Du, A.; Zhou, Z.; Lai, C.; Yu, W.; Yu, J.; *et al*. 2021. An atlas of bacterial secondary metabolite biosynthesis gene clusters. *Environmental Microbiology* 23: 6981–6992.

Wenski, S.L.; Thiengmag, S.; Helfrich, E.J.N. 2022. Complex peptide natural products: Biosynthetic principles, challenges and opportunities for pathway engineering. *Synthetic and Systems Biotechnology* 7: 631–647.

Wittmann, F.; Householder, J.E.; Piedade, M.T.F.; Schöngart, J.; Demarchi, L.O.; Quaresma, A.C.; *et al*. 2022. A review of the ecological and biogeographic differences of Amazonian floodplain forests. *Water* 14: 3360. doi: 10.3390/w14213360

Wohlleben, W.; Mast, Y.; Stegmann, E.; Ziemert, N. 2016. Antibiotic drug discovery. *Microbial Biotechnology* 9: 541–548.

Ye, R.; Xu, H.; Wan, C.; Peng, S.; Wang, L.; Xu, H.; *et al*. 2013. Antibacterial activity and mechanism of action of ε-poly-l-lysine. *Biochemical and Biophysical Research Communications* 439: 148–153.

Ye, Y.-Q.; Ye, M.-Q.; Zhang, X.-Y.; Huang, Y.-Z.; Zhou, Z.-Y.; Feng, Y.-J.; *et al*. 2024. Description of the first marine-isolated member of the under-represented phylum *Gemmatimonadota*, and the environmental distribution and ecogenomics of *Gaopeijiales* ord. nov. *MSystems* 9: e0053524.

Yen, W.-Y.; Stern, K.; Mishra, S.; Helminiak, L.; Sanchez-Vicente, S.; Kim, H.K. 2021. Virulence potential of *Rickettsia amblyommatis* for spotted fever pathogenesis in mice. *Pathogens and Disease* 79: ftab024.

Zdouc, M.M.; Blin, K.; Louwen, N.L.L.; Navarro, J.; Loureiro, C.; Bader, C.D.; *et al*. 2025. MIBiG 4.0: advancing biosynthetic gene cluster curation through global collaboration. *Nucleic Acids Research* 53: D678–D690.

Zeineldin, M.; Hicks, J.; Ward, H.J.; Wünschmann, A.; Camp, P.; Farrell, D.; *et al*. 2023. Complete genome sequence of *Candidatus Mycobacterium wuenschmannii*, a nontuberculous mycobacterium isolated from a captive population of Amazon milk frogs. *Microbiology Resource Announcements* 12: e0054723.

Zhou, M.; Liu, F.; Yang, X.; Jin, J.; Dong, X.; Zeng, K.-W.; *et al*. 2018. Bacillibactin and bacillomycin analogues with cytotoxicities against human cancer cell lines from marine *Bacillus* sp. PKU-MA00093 and PKU-MA00092. *Marine Drugs* 16: 22. doi: 10.3390/md1601002

RECEIVED: 13/01/2024
ACCEPTED: 03/07/2025
ASSOCIATE EDITOR: André Willerding
DATA AVAILABILITY: Genome data are publicly available on the NCBI website where they were collected. All additional data that support the analyses are available from the corresponding author [Eric de Lima Silva Marques] upon reasonable request.
AUTHOR CONTRIBUTIONS: Eric de Lima Silva Marques - conceptualization, data curation, formal analysis, methodology, validation, writing - original draft, writing - review & editing; **Adriana Barros de Cerqueira e Silva** - conceptualization, formal analysis, investigation, methodology, writing - original draft, writing - review & editing; **Mariana Barros de Cerqueira e Silva** - conceptualization, formal analysis, investigation, methodology, writing - original draft, writing - review & editing.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

SUPPLEMENTARY MATERIAL

Marques *et al*. Exploring the biotechnological potential of Amazonian microorganisms through *in silico* genome analysis for detection of biosynthetic gene clusters

Table S1. Similarity scores higher than 1.0 predicted by MIBiG natively in AntiSMASH software from Amazonian microbial genomes available in the NCBI database.

Code	Closest similar compound	Similarity score	Type
C2	bacillibactin	1.10	NRP
D2	bacillaene	1.68	NRP/PKS
D2	bacillomycin	1.67	NRP/PKS
D2	amylocyclicin	2.61	RiPP
L2	shinorine	1.44	NRP
L2	pseudospumigin	1.26	NRP/PKS
Q2	thioangucycline	1.18	PKS
O2	thalassospiramide	1.24	NRP/PKS

Table S2. Distribution of NRP biosynthetic gene clusters identified on BIG-SCAPE using the genomes from Amazonian microbes available in the NCBI database.

Code	NAPAA	NAPAA.NRP-metallophore.NRPS	NAPAA.NRPS-like	NRP-metallophore.NRPS	NRPS	NRPS.NRPS-like	NRPS-like	Total
B1	0	0	0	0	1	0	0	1
C2	0	0	0	1	0	0	0	1
D2	0	0	0	0	1	0	0	1
E1	1	0	0	0	2	0	1	4
E2	0	0	0	0	1	0	0	1
F1	1	0	0	0	0	0	0	1
F2	0	0	0	0	1	0	0	1
G2	0	0	0	1	1	0	0	2
H1	1	0	0	0	0	0	1	2
H2	0	0	0	0	5	0	3	8
I1	1	0	0	0	0	0	1	2
J2	1	0	0	1	0	0	1	3
K2	1	0	0	0	0	0	2	3
L1	0	0	0	0	0	1	0	1
L2	0	0	0	0	5	1	1	7
M2	0	0	0	0	1	0	1	2
N2	0	0	0	2	0	0	0	2
O2	3	0	0	0	0	0	0	3
P1	0	0	0	0	0	1	1	2
P2	1	0	1	1	1	0	1	5
Q1	1	0	0	0	0	0	0	1
Q2	3	1	0	0	2	0	0	6
T2	1	0	0	0	0	0	1	2
Total	**15**	**1**	**1**	**6**	**21**	**3**	**14**	**61**

Table S3. Distribution of NRP.PKS biosynthetic gene clusters identified on BIG-SCAPE using the genomes from Amazonian microbes available in the NCBI database.

Code	NRP-metallophore.NRPS. NRPS-like.T1PKS	NRPS.T1PKS	NRPS.T3PKS.trans AT-PKS	NRPS-like.T1PKS	Total
D2	0	0	1	0	1
J2	0	1	0	0	1
K1	0	0	0	1	1
L2	0	2	0	0	2
M2	0	2	0	0	2
N1	0	1	0	0	1
O2	1	2	0	0	3
Q2	1	1	0	1	3
Total	**2**	**9**	**1**	**2**	**14**

Table S4. Distribution of NRPS.other (red), NRPS.PKS.other (green) and NRPS.terpene (blue) biosynthetic gene clusters identified on BIG-SCAPE using the genomes from Amazonian microbes available in the NCBI database.

Code	NRPS.ectoine	NRPS-like. arylpolyene	Total	NRPS.betalactone.transAT-PKS	Total	NRPS-like. terpene	Total
D2	0	0	0	1	1	0	0
G2	1	0	1	0	0	0	0
K1	0	1	1	0	0	0	0
M2	0	0	0	0	0	1	1
R1	1	0	1	0	0	0	0
Total	**2**	**1**	**3**	**1**	**1**	**1**	**1**

Table S5. Distribution of NRPS.RiPP biosynthetic gene clusters identified on BIG-SCAPE using the genomes from Amazonian microbes available in the NCBI database. LAP = LAP.NRPS.NRPS-like.mycosporine-like; NAPAA = NAPAA.lanthipeptide-class-ii.proteusin.

Code	LAP	NAPAA	NRP-metallophore. NRPS.RiPP-like	NRPS.NRPS-like.thiopeptide	NRPS.RiPP-like	NRPS-like.RiPP-like.proteusin	NRPS-like.thiopeptide	Total
D2	0	0	1	0	0	0	0	1
G2	0	0	0	0	1	1	0	2
L2	1	1	0	1	0	0	0	3
N2	0	0	0	0	0	0	1	1
Total	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**7**

Table S6. Distribution of biosynthetic gene clusters identified as "other" on BIG-SCAPE using the genomes from Amazonian microbes available in the NCBI database. AAA = acyl_amino_acids; AAAH = acyl_amino_acids.hserlactone; AP = arylpolyene; HL = hserlactone.

Code	2dos	AAA	AAAH	AP	betalactone	ectoine	HL	indole	ladderane	NAGGN	NI-siderophore	other	phosphonate	resorcinol	Total
A1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
A2	0	0	0	1	1	0	0	0	0	0	1	0	0	0	3
B1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
B2	0	0	0	1	1	0	0	0	0	0	1	0	0	0	3
C2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
D1	0	0	0	4	0	0	0	0	0	0	0	0	0	1	5
D2	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
E1	0	1	0	0	1	0	0	0	0	0	1	0	0	0	3
E2	0	0	0	0	1	0	3	0	0	0	0	0	0	0	4

Table S6. Continued

Code	2dos	AAA	AAAH	AP	betalactone	ectoine	HL	indole	ladderane	NAGGN	NI-siderophore	other	phosphonate	resorcinol	Total
F1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	2
F2	0	0	0	0	1	0	3	0	0	0	0	0	0	0	4
G1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
G2	0	0	0	0	0	0	1	0	0	0	0	0	1	0	2
H1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	2
H2	0	0	0	0	0	0	2	0	0	0	0	0	2	0	4
I1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	2
J1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	2
K1	0	0	0	1	0	1	1	0	1	0	0	0	0	0	4
L2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
M1	0	0	0	2	0	0	0	0	0	0	0	0	0	0	2
N1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	2
N2	0	0	0	1	0	0	0	0	0	1	0	0	0	0	2
O1	0	0	0	0	0	1	3	0	0	0	0	0	0	0	4
O2	1	0	0	0	0	1	0	0	0	0	0	0	0	0	2
P1	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2
P2	1	0	0	0	1	1	0	0	0	0	0	0	1	0	4
Q2	1	0	0	0	2	2	0	0	0	0	0	0	0	0	5
R2	0	0	0	1	1	0	0	0	0	0	1	0	0	0	3
S1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
T2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
Total	**3**	**1**	**1**	**13**	**14**	**7**	**17**	**1**	**2**	**1**	**8**	**1**	**4**	**1**	**74**

Table S7. Distribution of PKS (red), PKS.other (green) and PKS.RiPP (blue) biosynthetic gene clusters identified on BIG-SCAPE using the genomes from Amazonian microbes available in the NCBI database.

Code	hglE-KS	PKS-like	T1PKS	T1PKS.hglE-KS	T3PKS	transAT-PKS	Total	T3PKS.betalactone	Total	RiPP-like.T3PKS	Total
B1	0	0	1	0	0	0	1	0	0	0	0
D2	0	1	0	0	1	2	4	0	0	0	0
J1	0	0	0	0	1	0	1	0	0	0	0
J2	0	0	6	0	1	0	7	1	1	0	0
K2	0	0	0	0	0	0	0	0	0	1	1
L2	1	0	1	0	0	0	2	0	0	0	0
M2	0	0	0	1	0	0	1	0	0	0	0
P1	0	0	0	0	1	0	1	0	0	0	0
P2	1	1	2	0	0	0	4	0	0	0	0
Q1	0	0	0	0	1	0	1	0	0	0	0
Q2	0	1	3	0	0	0	4	0	0	0	0
T2	0	0	0	0	1	0	1	0	0	0	0
Total	**2**	**3**	**13**	**1**	**6**	**2**	**27**	**1**	**1**	**1**	**1**

Table S8. Distribution of NRPS.RiPP biosynthetic gene clusters identified on BIG-SCAPE using the genomes from Amazonian microbes available in the NCBI database. CLALC = cyclic-lactone-autoinducer.lanthipeptide-class-ii; LPii = lanthipeptide-class-ii; LPiii = lanthipeptide-class-iii; LPv = lanthipeptide-class-v; LAP.T = LAP.thiopeptide; RC = redox-cofactor; RRE = RRE-containing; RRE-CLvS = RRE-containing.lanthipeptide-class-v.spliceotide; RRE-CL = RRE-containing.lassopeptide.

Code	CLALC	LPii	LPiii	LPv	LAP	LAP.T	lassopeptide	microviridin	ranthipeptide	RC	RiPP-like	RRE	RRE-CLvS	RRE-CL	Total
A1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	2
A2	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
B1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	2
B2	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
C1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	2
C2	1	1	0	0	1	0	0	0	0	0	3	0	0	0	6
D1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
D2	0	1	1	0	0	0	0	0	0	0	0	0	0	0	2
E1	0	0	0	0	0	0	0	0	0	1	3	0	0	0	4
E2	0	0	0	0	0	0	0	0	0	1	2	0	0	0	3
F1	0	0	0	0	0	0	0	0	0	0	0	2	0	1	3
F2	0	0	0	0	0	0	0	0	0	1	2	0	0	0	3
G2	0	0	0	0	0	0	0	0	0	1	0	1	0	0	2
H1	0	0	0	0	0	0	0	0	0	2	1	1	0	0	4
H2	0	0	0	0	0	0	0	0	0	1	1	0	0	0	2
I1	0	0	0	0	0	0	0	0	0	2	1	1	0	0	4
I2	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
J1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
J2	0	0	0	0	0	0	0	0	0	1	1	0	0	0	2
K1	0	0	0	0	0	0	0	0	0	2	2	0	0	0	4
K2	0	0	0	0	0	0	0	0	0	0	1	2	0	0	3
L1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	2
L2	0	1	0	3	2	0	0	2	0	0	1	0	1	1	11
M1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
M2	0	1	0	2	0	1	0	1	0	0	2	0	0	1	8
N1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	2
N2	0	0	0	0	0	0	0	0	1	1	2	0	0	0	4
O1	0	0	0	0	0	0	4	0	0	0	1	0	0	0	5
O2	0	0	0	0	0	0	1	0	1	1	3	0	0	0	6
P1	0	0	0	1	0	0	0	0	0	0	2	3	0	0	6
P2	0	0	0	0	0	0	0	0	1	1	3	0	0	0	5
Q1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Q2	0	0	0	0	0	0	0	0	2	1	5	0	0	0	8
R1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
R2	0	0	0	0	0	0	0	0	0	1	1	0	0	0	2
Total	**1**	**5**	**1**	**6**	**3**	**1**	**7**	**3**	**5**	**19**	**45**	**13**	**1**	**4**	**114**

Table S9. Distribution of RiPP.other (red), oligosaccharide (green) and terpene (blue) biosynthetic gene clusters identified as on BIG-SCAPE using the genomes from Amazonian microbes available in the NCBI database.

Code	indole.thioamitides .thiopeptide	Lassopeptide .resorcinol	Total	oligosaccharide	Total	terpene	Total
A1	0	0	0	0	0	2	2
B2	0	0	0	0	0	1	1
C1	0	0	0	0	0	1	1
C2	0	0	0	0	0	1	1
D1	0	0	0	0	0	1	1
D2	0	0	0	0	0	2	2
E1	0	0	0	0	0	3	3
E2	0	0	0	0	0	2	2
F1	0	0	0	0	0	1	1
F2	0	0	0	0	0	2	2
G1	0	0	0	0	0	1	1
G2	1	0	1	0	0	2	2
H2	0	0	0	0	0	2	2
I2	0	0	0	0	0	2	2
J1	0	0	0	0	0	1	1
J2	0	0	0	0	0	1	1
K1	0	0	0	0	0	3	3
K2	0	0	0	0	0	1	1
L1	0	0	0	0	0	1	1
L2	0	0	0	0	0	4	4
M1	0	0	0	0	0	1	1
M2	0	0	0	0	0	2	2
O1	0	0	0	0	0	1	1
O2	0	0	0	1	1	2	2
P1	0	1	1	0	0	2	2
P2	0	0	0	0	0	3	3
Q1	0	0	0	0	0	1	1
Q2	0	0	0	0	0	6	6
R1	0	0	0	0	0	2	2
R2	0	0	0	0	0	1	1
T2	0	0	0	0	0	1	1
U2	0	0	0	0	0	3	3
Total	**1**	**1**	**2**	**1**	**1**	**59**	**59**

Table S10. Categories of biosynthetic gene clusters identified on BIG-SCAPE using the genomes from Amazonian microbes available in the NCBI database. .O = other.

Code	NRPS	NRPS.O	NRPS. PKS	NRPS. PKS.O	NRPS. RiPP	NRPS. terpene	other	PKS	PKS.O	PKS. RiPP	RiPP	RiPP.O	saccharide	terpene	Total
A1	0	0	0	0	0	0	1	0	0	0	2	0	0	2	5
A2	0	0	0	0	0	0	3	0	0	0	1	0	0	0	4
B1	1	0	0	0	0	0	1	1	0	0	2	0	0	0	5
B2	0	0	0	0	0	0	3	0	0	0	1	0	0	1	5
C1	0	0	0	0	0	0	0	0	0	0	2	0	0	1	3

Table S10. Continued

Code	NRPS	NRPS.O	NRPS. PKS	NRPS. PKS.O	NRPS. RiPP	NRPS. terpene	other	PKS	PKS.O	PKS. RiPP	RiPP	RiPP.O	saccharide	terpene	Total
C2	1	0	0	0	0	0	1	0	0	0	6	0	0	1	9
D1	0	0	0	0	0	0	5	0	0	0	1	0	0	1	7
D2	1	0	1	1	1	0	1	4	0	0	2	0	0	2	13
E1	4	0	0	0	0	0	3	0	0	0	4	0	0	3	14
E2	1	0	0	0	0	0	4	0	0	0	3	0	0	2	10
F1	1	0	0	0	0	0	2	0	0	0	3	0	0	1	7
F2	1	0	0	0	0	0	4	0	0	0	3	0	0	2	10
G1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	2
G2	2	1	0	0	2	0	2	0	0	0	2	1	0	2	12
H1	2	0	0	0	0	0	2	0	0	0	4	0	0	0	8
H2	8	0	0	0	0	0	4	0	0	0	2	0	0	2	16
I1	2	0	0	0	0	0	2	0	0	0	4	0	0	0	8
I2	0	0	0	0	0	0	0	0	0	0	1	0	0	2	3
J1	0	0	0	0	0	0	2	1	0	0	1	0	0	1	5
J2	3	0	1	0	0	0	0	7	1	0	2	0	0	1	15
K1	0	1	1	0	0	0	4	0	0	0	4	0	0	3	13
K2	3	0	0	0	0	0	0	0	0	1	3	0	0	1	8
L1	1	0	0	0	0	0	0	0	0	0	2	0	0	1	4
L2	7	0	2	0	3	0	1	2	0	0	11	0	0	4	30
M1	0	0	0	0	0	0	2	0	0	0	1	0	0	1	4
M2	2	0	2	0	0	1	0	1	0	0	8	0	0	2	16
N1	0	0	1	0	0	0	2	0	0	0	2	0	0	0	5
N2	2	0	0	0	1	0	2	0	0	0	4	0	0	0	9
O1	0	0	0	0	0	0	4	0	0	0	5	0	0	1	10
O2	3	0	3	0	0	0	2	0	0	0	6	0	1	2	17
P1	2	0	0	0	0	0	2	1	0	0	6	1	0	2	14
P2	5	0	0	0	0	0	4	4	0	0	5	0	0	3	21
Q1	1	0	0	0	0	0	0	1	0	0	0	0	0	1	3
Q2	6	0	3	0	0	0	5	4	0	0	8	0	0	6	32
R1	0	1	0	0	0	0	0	0	0	0	1	0	0	2	4
R2	0	0	0	0	0	0	3	0	0	0	2	0	0	1	6
S1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
T2	2	0	0	0	0	0	1	1	0	0	0	0	0	1	5
U2	0	0	0	0	0	0	0	0	0	0	0	0	0	3	3
Total	**61**	**3**	**14**	**1**	**7**	**1**	**74**	**27**	**1**	**1**	**114**	**2**	**1**	**59**	**366**

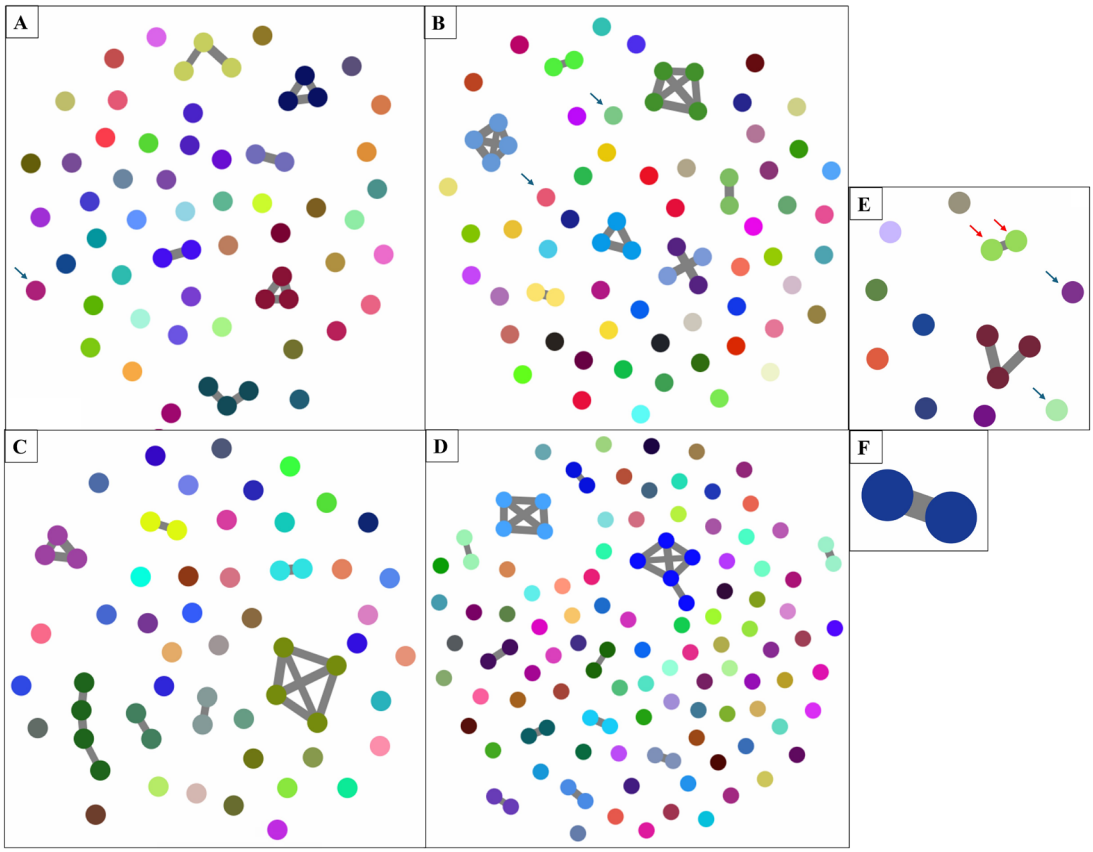
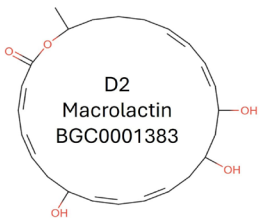


Figure S1. - Network layout generated by BiG-SCAPE for BGC classes identified in the analyzed genomes: NRP (**A**), other (**B**), terpene (**C**), RiPP (**D**), NRPS-PKS hybrid (**E**), and saccharide (**F**). Nodes represent individual biosynthetic gene clusters (BGCs); edges indicate similarity relationships based on domain architecture and sequence. Node colors correspond to different gene cluster families. Disconnected nodes represent singletons. Classes not shown consisted exclusively of singletons. The blue arrow indicates nodes associated with BGCs matched to MiBIG entries, whereas the remaining BGCs are linked exclusively to singleton classes. The red arrow highlights BGCs associated with attinimicin (Figure S14)

NZ_CP040378.1.region006: Bacillus velezensis strain P34 chromosome
BGC0000181.1: Macrolactin biosynthetic gene cluster
BGC0001383.1: Macrolactin biosynthetic gene cluster

NZ_CP040378 ↔ BGC0000181
ctg1_1395 ↔ CAJ57405.1 : 100.0%
ctg1_1386 ↔ CAG23963.1 : 98.6%
ctg1_1387 ↔ CAG23964.1 : 98.5%
ctg1_1393 ↔ CAG23970.2 : 98.4%
ctg1_1388 ↔ CAG23965.1 : 98.2%
ctg1_1389 ↔ CAG23966.1 : 98.1%
ctg1_1390 ↔ CAG23967.1 : 98.0%
ctg1_1392 ↔ CAG23969.1 : 97.8%
ctg1_1391 ↔ CAG23968.1 : 97.3%
ctg1_1394 ↔ CAJ57404.1 : 94.9%



NZ_CP040378 ↔ BGC0001383
ctg1_1394 ↔ AIJ04688.2 : 98.9%
ctg1_1386 ↔ AIJ04682.1 : 98.7%
ctg1_1392 ↔ AIJ04685.1 : 98.2%
ctg1_1393 ↔ AIJ04687.1 : 98.2%
ctg1_1390 ↔ AIJ04684.1 : 98.2%
ctg1_1387 ↔ AIJ04681.1 : 98.1%
ctg1_1389 ↔ AIJ04683.1 : 97.9%
ctg1_1391 ↔ AIJ04686.1 : 97.6%
ctg1_1388 ↔ AIJ04680.1 : 97.5%

Figure S2. Comparison of the macrolactin biosynthetic clusters (BGC0000181 and BGC0001383) with identified cluster in *Bacillus velezensis* (NZ_CP040378). Similarities between corresponding genes in the aligned regions are shown. The chemical structure of the closely related biosynthetic product is also presented. This BGC is classified as a polyketide.

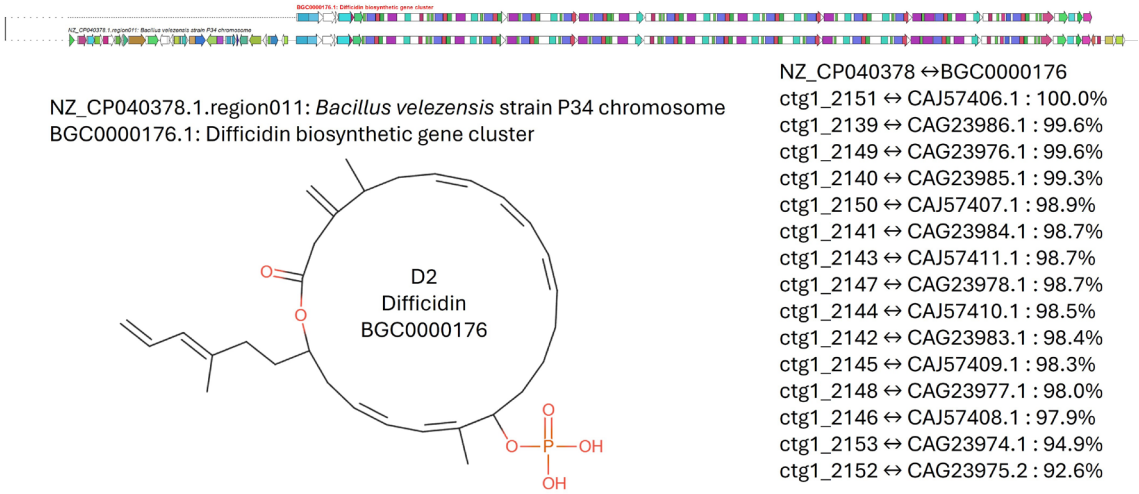


Figure S3. Comparison of the difficidin biosynthetic cluster (BGC0000176) with an identified cluster in *Bacillus velezensis* (NZ_CP040378). Similarities between corresponding genes in the aligned regions are shown. The chemical structure of the closely related biosynthetic product is also presented. This BGC is classified as a polyketide.

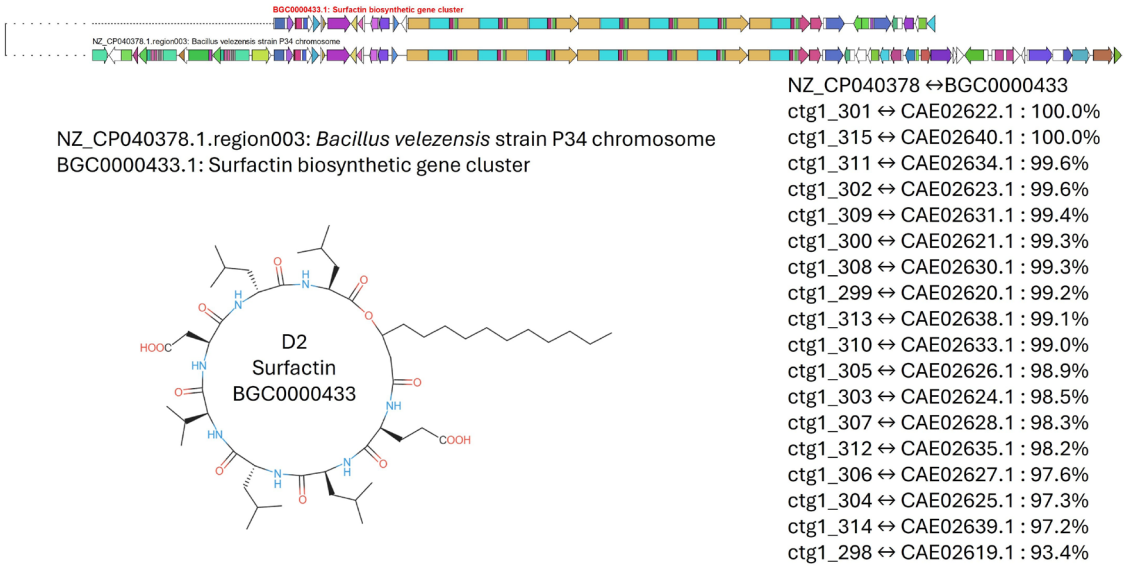


Figure S4. Comparison of the surfactin biosynthetic cluster (BGC0000433) with an identified cluster in *Bacillus velezensis* (NZ_CP040378). Similarities between corresponding genes in the aligned regions are shown. The chemical structure of the closely biosynthetic product is also presented. This BGC is classified as a NRPS.

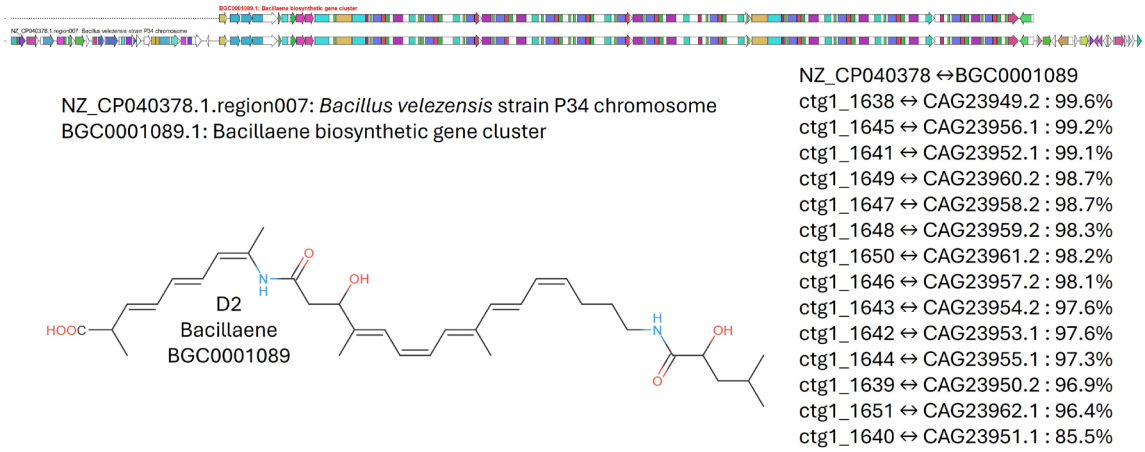


Figure S5. Comparison of the bacillaene biosynthetic cluster (BGC0001089) with an identified cluster in *Bacillus velezensis* (NZ_CP040378). Similarities between corresponding genes in the aligned regions are shown. The chemical structure of the closely related biosynthetic product is also presented. This BGC is classified as a NRPS.PKS.

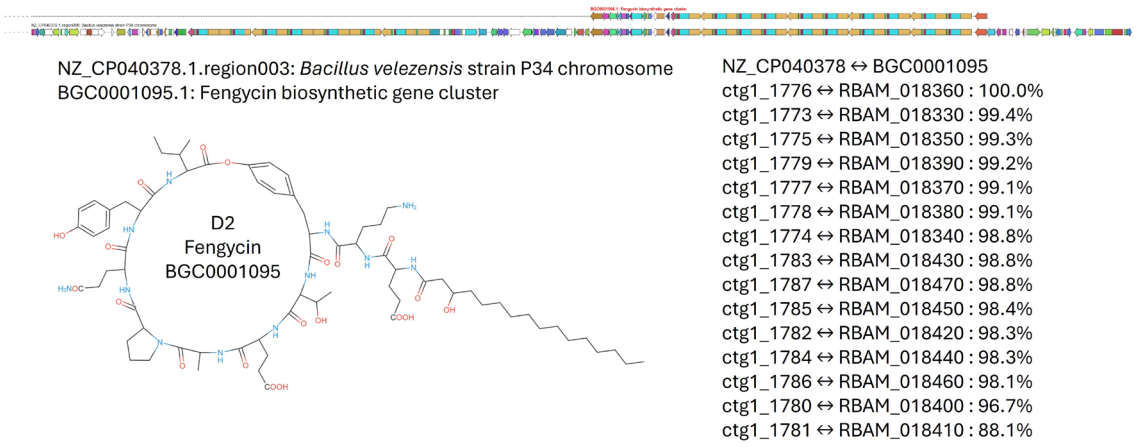


Figure S6. Comparison of the fengycin biosynthetic cluster (BGC0001095) with an identified cluster in *Bacillus velezensis* (NZ_CP040378). Similarities between corresponding genes in the aligned regions are shown. The chemical structure of the closely biosynthetic product is also presented. This BGC is classified as a NRPS.

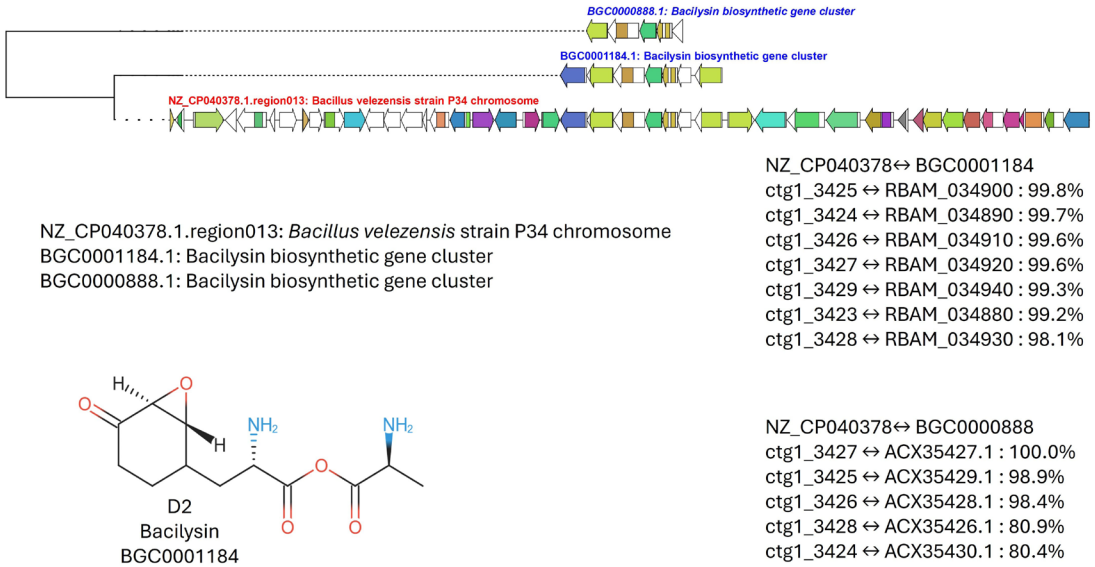


Figure S7. Comparison of the bacilysin biosynthetic clusters (BGC0000888 and BGC0001184) with an identified cluster in *Bacillus velezensis* (NZ_CP040378). Similarities between corresponding genes in the aligned regions are shown. The chemical structure of the closely related biosynthetic product is also presented. This BGC is classified as "other".

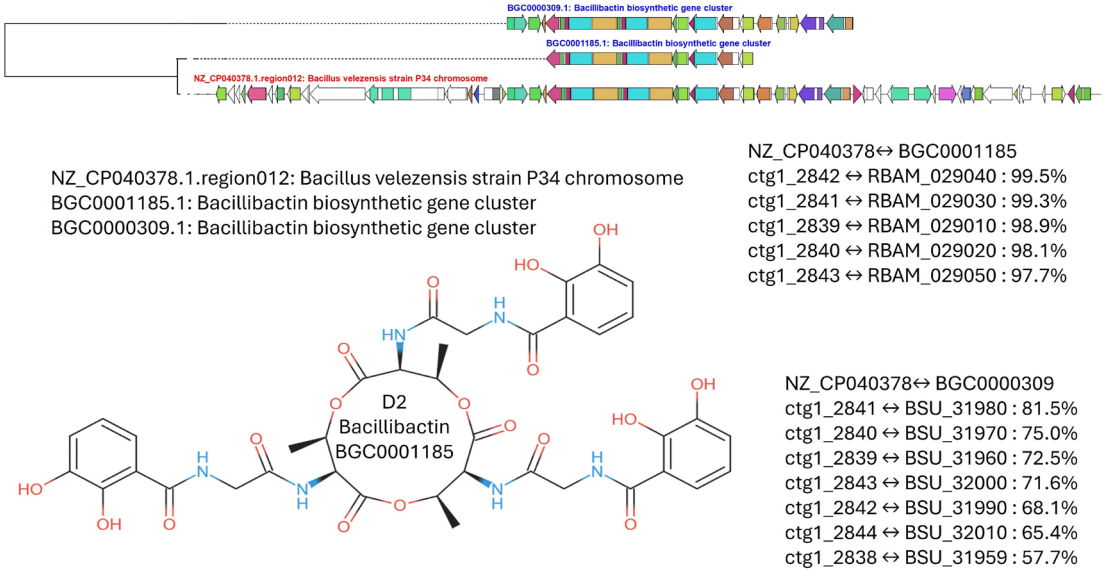


Figure S8. Comparison of the bacillibactin biosynthetic clusters (BGC0000309 and BGC0001185) with an identified cluster in *Bacillus velezensis* (NZ_CP040378). Similarities between corresponding genes in the aligned regions are shown. The chemical structure of the closely related biosynthetic product is also presented. This BGC is classified as a NRPS.RiPP.

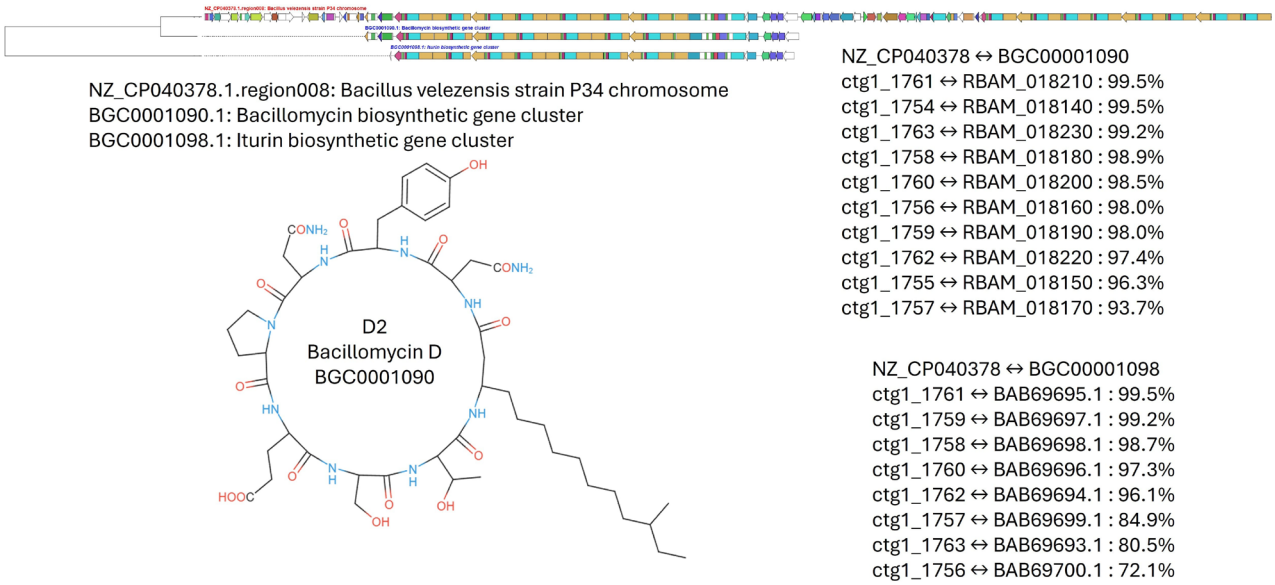
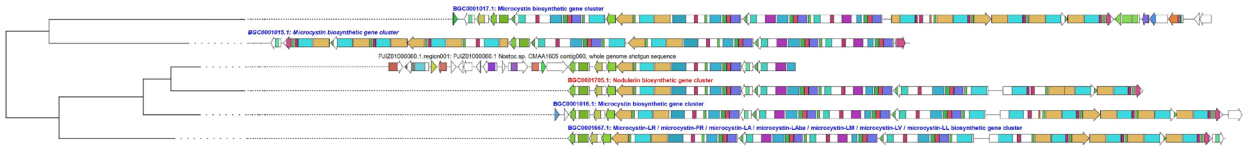
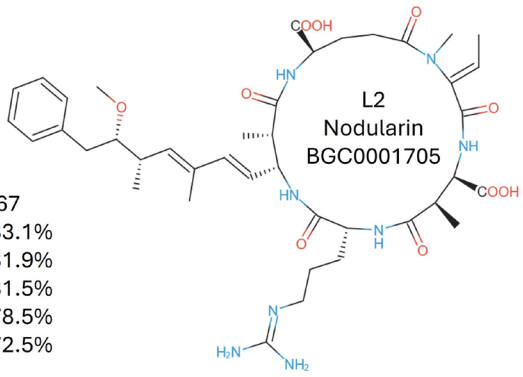


Figure S9. Comparison of the bacillomycin biosynthetic cluster (BGC0001090) and the iturin biosynthetic cluster (BGC0001098) with an identified cluster in *Bacillus velezensis* (NZ_CP040378). Similarities between corresponding genes in the aligned regions are shown. The chemical structure of the closely related biosynthetic product is also presented. These BGCs are classified as NRPS.PKS.other.



PJIZ01000060.1.region001: PJIZ01000060.1 Nostoc sp. CMAA1605
BGC0001705.1: Nodularin biosynthetic gene cluster
BGC0001016.1: Microcystin biosynthetic gene cluster
BGC0001667.1: Microcystin-LR /-FR /-LA /-LAba /-LM /-LV /-LL biosynt. gene cluster

PJIZ01000060↔ BGC0001705
ctg60_23 ↔ ATP76240.1 : 96.4%
ctg60_20 ↔ ATP76237.1 : 95.8%
ctg60_22 ↔ ATP76239.1 : 94.8%
ctg60_21 ↔ ATP76238.1 : 94.7%
ctg60_19 ↔ ATP76236.1 : 94.1%



PJIZ01000060↔ BGC0001016
ctg60_23 ↔ AAO62583.1 : 95.2%
ctg60_22 ↔ AAO62582.1 : 93.8%
ctg60_20 ↔ AAO62580.1 : 92.9%
ctg60_19 ↔ AAO62579.1 : 90.3%
ctg60_21 ↔ AAO62581.1 : 89.4%

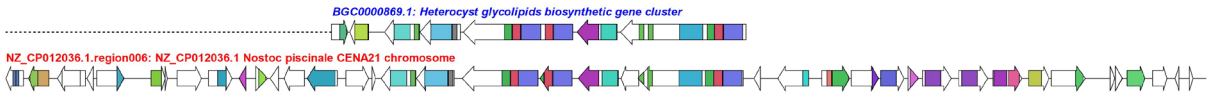
PJIZ01000060↔ BGC0001667
ctg60_23 ↔ AQH32487.1 : 83.1%
ctg60_20 ↔ AQH32479.1 : 81.9%
ctg60_22 ↔ AQH32481.1 : 81.5%
ctg60_19 ↔ AQH32478.1 : 78.5%
ctg60_21 ↔ AQH32480.1 : 72.5%

Figure S10. Comparison of the nodularin biosynthetic cluster BGC0001705 and the microcystin biosynthetic clusters (BGC0001016 and BGC0001667) with an identified cluster in *Nostoc sp.* CMA1605 (PJIZ01000060). Similarities between corresponding genes in the aligned regions are shown. The chemical structure of the closely related biosynthetic product is also presented. These BGCs are classified as NRPS.PKS.

PJIZ01000057.1.region001: PJIZ01000057.1 Nostoc sp. CMAA1605
BGC0001748.1: Pseudospumigin A / B / C / D / E / F biosynthetic gene cluster

PJIZ01000057↔ BGC0001748
ctg57_17 ↔ ATP76247.1 : 95.1%
ctg57_21 ↔ ATP76245.1 : 90.8%
ctg57_20 ↔ ATP76246.1 : 89.4%

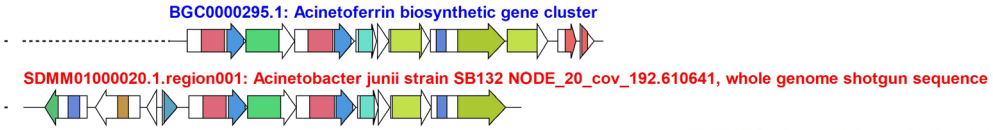
Figure S11. Comparison of the pseudospumigin biosynthetic cluster BGC0001748 and with an identified cluster in *Nostoc sp.* CMA1605 (PJIZ01000057). Similarities between corresponding genes in the aligned regions are shown. The chemical structure of the closely related biosynthetic product is not available. This BGC is classified as a NRPS.RiPP.



NZ_CP012036.1.region006: NZ_CP012036.1 Nostoc piscinale CENA21
BGC0000869.1: Heterocyst glycolipids biosynthetic gene cluster

NZ_CP012036↔ BGC0000869
ctg1_1681 ↔ AGJ76605.1 : 78.6%
ctg1_1682 ↔ AGJ76604.1 : 76.8%
ctg1_1685 ↔ AGJ76602.1 : 69.8%
ctg1_1687 ↔ AGJ76601.1 : 65.6%

Figure S12. Comparison of the heterocyst glycolipids biosynthetic cluster BGC0000869 with an identified cluster in *Nostoc piscinale* (NZ_CP012036). Similarities between corresponding genes in the aligned regions are shown. The chemical structure of the closely related biosynthetic product is not available. This BGC is classified as a polyketide.



SDMM01000020.1.region001: *Acinetobacter junii* strain SB132
BGC0000295.1: Acinetoferrin biosynthetic gene cluster

SDMM01000020↔ BGC0000295
ctg12_123 ↔ BAM09328.1 : 97.4%
ctg12_124 ↔ BAM09329.1 : 97.2%
ctg12_122 ↔ BAM09327.1 : 95.9%
ctg12_126 ↔ BAM09331.1 : 93.7%
ctg12_125 ↔ BAM09330.1 : 92.7%
ctg12_121 ↔ BAM09326.1 : 91.9%

Figure S13. Comparison of the acinetoferrin biosynthetic cluster BGC0000295 with an identified cluster in *Acinetobacter junii* (SDMM01000020). Similarities between corresponding genes in the aligned regions are shown. The chemical structure of the closely related biosynthetic product is not available. This BGC is classified as "other".

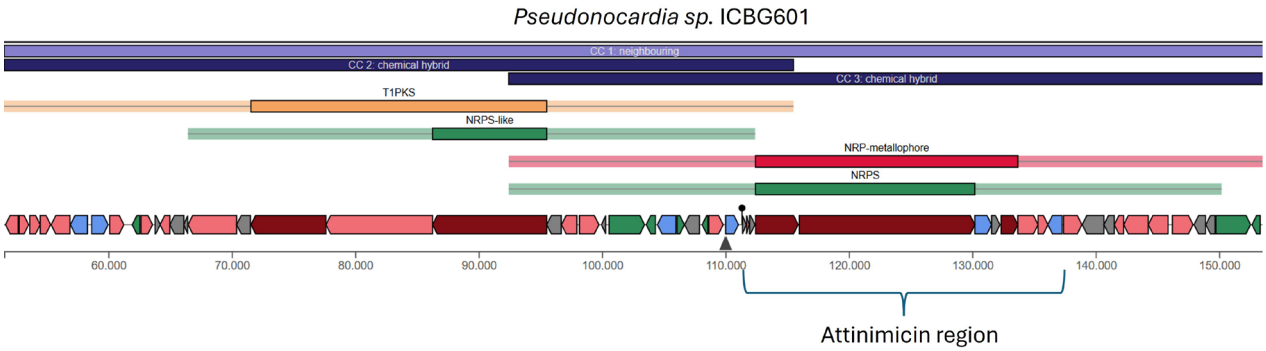


Figure S14. Putative biosynthetic gene cluster from attinimicin proposed by Fukuda *et al.* 2021 and identified at *Pseudonocardia* sp. ICBG601. The biosynthetic gene cluster was also identified at *Pseudonocardia sp.* ICBG1034 through BIG-SCAPE.