



OPEN ACCESS

EDITED BY

Fidele Tugizimana,
University of Johannesburg, South Africa

REVIEWED BY

Renu Pandey,
Institute of Bioinformatics (IOB), India
Emanuela Camera,
San Gallicano Dermatological Institute -
IRCCS, Italy

*CORRESPONDENCE

Ricardo Roberto da Silva,
✉ ridasilva@usp.br

†These authors have contributed equally
to this work

RECEIVED 30 October 2024

ACCEPTED 17 February 2025

PUBLISHED 20 March 2025

CITATION

Arini GS, Borelli TC, Ferreira EG, de Felício R, Rezende-Teixeira P, Pedrino M, Rabiço F, de Siqueira GMV, Mencucini LG, Tsuji H, Neves Andrade LS, Garrido LM, Padilla G, Gil-de-la-Fuente A, Wang M, Lopes NP, Barbosa Trivella DB, Costa-Lotufo LV, Guazzaroni M-E and Roberto da Silva R (2025) A multi-omics reciprocal analysis for characterization of bacterial metabolism. *Front. Mol. Biosci.* 12:1515276. doi: 10.3389/fmolb.2025.1515276

COPYRIGHT

© 2025 Arini, Borelli, Ferreira, de Felício, Rezende-Teixeira, Pedrino, Rabiço, de Siqueira, Mencucini, Tsuji, Neves Andrade, Garrido, Padilla, Gil-de-la-Fuente, Wang, Lopes, Barbosa Trivella, Costa-Lotufo, Guazzaroni and Roberto da Silva. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A multi-omics reciprocal analysis for characterization of bacterial metabolism

Gabriel Santos Arini^{1,2,3†}, Tiago Cabral Borelli^{1,2,3†}, Elthon Góis Ferreira⁴, Rafael de Felício⁵, Paula Rezende-Teixeira⁴, Matheus Pedrino^{3,6}, Franciene Rabiço^{3,6}, Guilherme Marcelino Viana de Siqueira^{3,6}, Luiz Gabriel Mencucini^{1,2}, Henrique Tsuji^{1,2}, Lucas Sousa Neves Andrade⁷, Leandro Maza Garrido⁷, Gabriel Padilla⁷, Alberto Gil-de-la-Fuente^{8,9}, Mingxun Wang¹⁰, Norberto Peporine Lopes², Daniela Barretto Barbosa Trivella⁵, Leticia Veras Costa-Lotufo⁴, María-Eugenia Guazzaroni⁶ and Ricardo Roberto da Silva^{1,2,3*}

¹Computational Chemical Biology Laboratory, Department of BioMolecular Sciences, School of Pharmaceutical Sciences of Ribeirão Preto, University of São Paulo, São Paulo, Brazil, ²NPPNS, Department of BioMolecular Sciences, School of Pharmaceutical Sciences of Ribeirão Preto, University of São Paulo, São Paulo, Brazil, ³Cellular and Molecular Biology Program, Department of Cellular and Molecular Biology of Ribeirão Preto, School of Medicine, University of São Paulo, São Paulo, Brazil, ⁴Marine Pharmacology Laboratory, Department of Pharmacology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil, ⁵Brazilian Biosciences National Laboratory (LNBio), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas, Brazil, ⁶MetaGenLab Laboratory, Department of Biology, FFCLRP, University of São Paulo of Ribeirão Preto, School of Medicine, University of São Paulo, São Paulo, Brazil, ⁷Laboratory of Bioproducts, Department of Microbiology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil, ⁸Centro de Metabolómica y Bioanálisis (CEMBIO), Facultad de Farmacia, Universidad San Pablo-CEU, CEU Universities, Urbanización Montepríncipe, Boadilla del Monte, Spain, ⁹Departamento de Tecnologías de la Información, Escuela Politécnica Superior, Universidad San Pablo-CEU, CEU Universities, Urbanización Montepríncipe, Boadilla del Monte, Spain, ¹⁰Department of Computer Science and Engineering, University of California Riverside, Riverside, CA, United States

Introduction: Exploiting microbial natural products is a key pursuit of the bioactive compound discovery field. Recent advances in modern analytical techniques have increased the volume of microbial genomes and their encoded biosynthetic products measured by mass spectrometry-based metabolomics. However, connecting multi-omics data to uncover metabolic processes of interest is still challenging. This results in a large portion of genes and metabolites remaining unannotated. Further exacerbating the annotation challenge, databases and tools for annotation and omics integration are scattered, requiring complex computations to annotate and integrate omics datasets.

Methods: Here we performed a two-way integrative analysis combining genomics and metabolomics data to describe a new approach to characterize the marine bacterial isolate BRA006 and to explore its biosynthetic gene cluster (BGC) content as well as the bioactive compounds detected by metabolomics.

Results and Discussion: We described BRA006 genomic content and structure by comparing Illumina and Oxford Nanopore MinION sequencing approaches. Digital DNA:DNA hybridization (dDDH) taxonomically assigned BRA006 as a potential new species of the *Micromonospora* genus.

Starting from LC-ESI(+)-HRMS/MS data, and mapping the annotated enzymes and metabolites belonging to the same pathways, our integrative analysis allowed us to correlate the compound Brevianamide F to a new BGC, previously assigned to other function.

KEYWORDS

untargeted metabolomics, genomics, multi-omics analysis, bioinformatics, microbiology

1 Introduction

The search for new bioactive compounds of natural origin from different organisms coming from different biomes is an arduous task. Since marine environments are poorly explored, they hold the promise of a formidable source of rich metabolic potential for the production of novel biosynthetic compounds, especially when you consider the microorganisms that reach a billion strains in a Gram of marine sediment (Fenical and Jensen, 2006). Considering that more than 40% of pharmaceutical ingredients are derived directly or indirectly from natural products derived from plants or microorganisms, one can expect that thousands of unknown potential medicines are expected to be discovered in marine ecosystems (Newman and Cragg, 2020; Kim et al., 2021). This feature places Brazil under the spotlight since its coast is especially large, ranging from tropical to temperate climate zones (Wilke et al., 2021).

The *Micromonospora* genus is composed by 177 Gram-positive, spore-forming aerobic species found mainly in marine environments. Also, this genus belongs to the phylum Actinomycetota, which is responsible for 70% of natural compounds under development or already in clinical use (Hifnawy et al., 2020). The chemical diversity, in terms of natural products, that this genus is capable of producing is enormous. *Micromonospora* natural products are used as drugs against infections caused by fungi or bacteria. The genera started to receive attention after the discovery of gentamicin in 1963 and after that, more than 740 bioactive compounds have been reported from *Micromonospora* strains. Among this chemical diversity produced, as well as the different locations where this genus can be found, there are reports in the literature searching specifically for anticancer bioactive compounds on *Micromonospora* sp. BRA006 (Sousa et al., 2012).

The evolution of bacterial genome organization clustered genes that encode enzymes of the same metabolic pathway, which are known as biosynthetic gene clusters (BGC) (Ruzzini and Clardy, 2016). For this reason, modern drug discovery from bacteria is based on BGC identification as the starting point followed by an experimental procedure that aims to detect, isolate, or produce the compound (Domingues Vieira et al., 2022). However, searching for novel natural bioactive compounds from microorganisms can be a harsh task, mostly because the majority of the microbial life cannot be cultured under laboratory conditions. Also, it is difficult to obtain bioactive compounds in the desired concentrations (Sun et al., 2019). Thus, bacterial bioactive compound discovery requires a multidisciplinary approach, such as genomics and metabolomics (Behsaz et al., 2021). Genomics enables the analysis of the whole genome sequencing data and raises hypotheses about metabolic pathways and compound products based on the genetic content (Andrade, 2020). Then, metabolomic assays based on

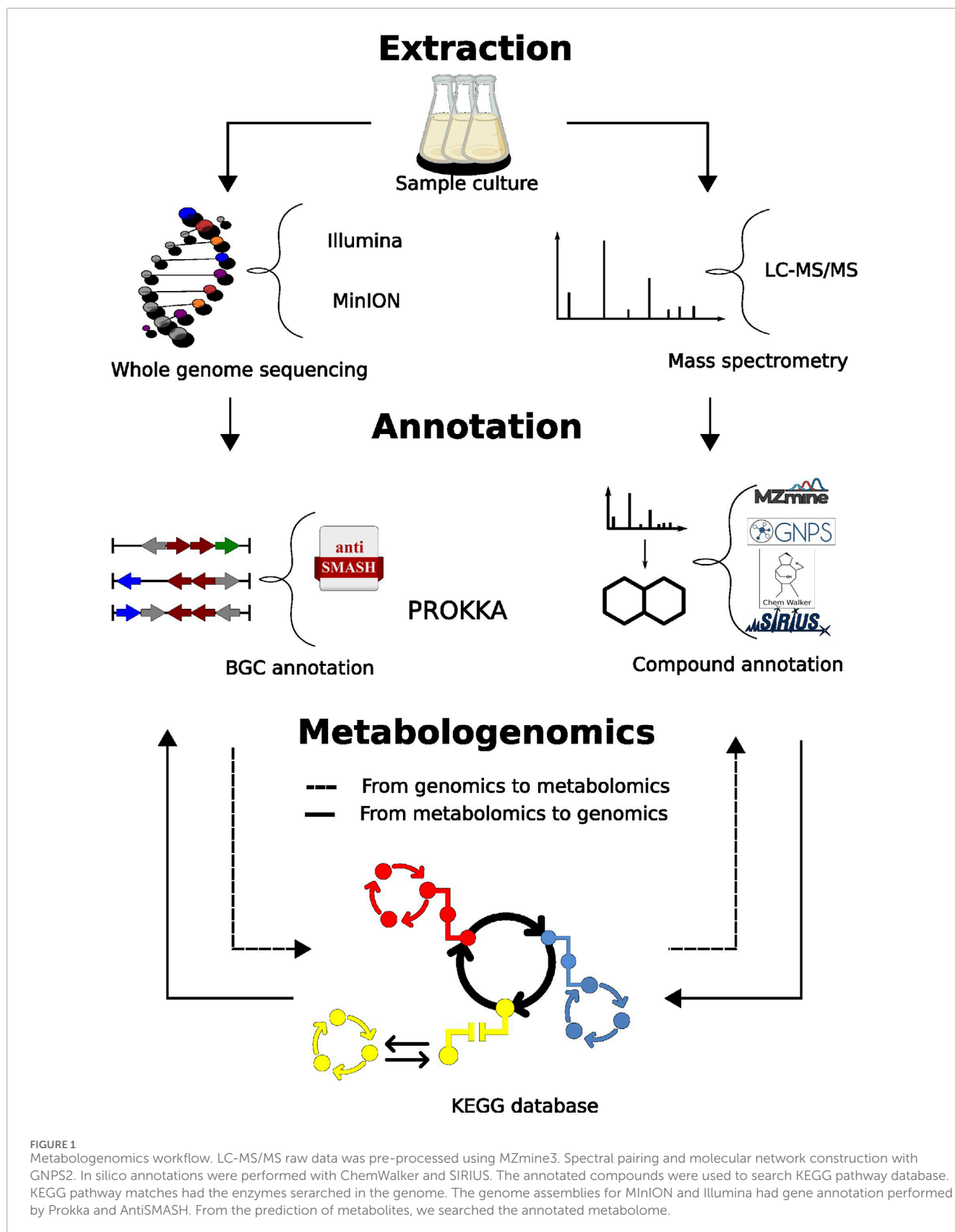
mass spectrometry analysis, such as LC-MS/MS are performed to validate them as well as be an important step in order to identify new possible bioactive compounds. The integration of these two omics sciences through multi-omics approaches opens up the possibility of accessing how much of a given compound predicted in a BGC is actually being produced and vice versa. And also to search for new BGCs starting from the annotated metabolites. Additionally, data integration has shown a promising resource in the description of new bacterial strains, as it allows the study and characterization of new microorganisms in a holistic way (Kato et al., 2024).

In the present work, we used multi-omics approach to describe the potential bioactive compounds of BRA006 (Figure 1), a bacteria strain recovered from a marine environment collected on the coast of Brazil. We performed our analysis in a two-way direction, searching for metabolites by LC-MS/MS previously predicted by antiSMASH, using two genome sequencing platforms, as well as finding coding sequences (CDS) for enzymes that are part of metabolic pathways for syntheses of BRA006s observed metabolome.

2 Materials and methods

2.1 Collection, DNA extraction, and sequencing

The BRA006, from MicroMarin collection (<https://www.labbmar.ufc.br/micromarinbr>) was cultured in A1 medium [Starch (10 g/L); Yeast extract (4 g/L); Peptone (2 g/L); Sea Water 75%] in a volume of 100 mL in 250 mL Erlenmeyer flasks. The cultures were centrifuged at $12,000 \times g$ for 10 min and the cell pellets were resuspended in 10 μ L of lysozyme 0.05 g/mL for 500 μ L of SET buffer. The mixture was incubated at 37°C for 30 min. Then 14 μ L Proteinase K (20 mg/mL) and 60 μ L SDS 10% were added to the cell lysate and incubated for a further 1 h at 55°C. Then 200 μ L NaCl 5 M was added and the temperature was raised to 37°C. 500 μ L of chloroform was added and the system was centrifuged at $4,500 \times g$ for 10 min at 20°C. 500 μ L was collected in a new tube and 300 μ L of isopropanol was added and incubated overnight (16 h). The system was centrifuged at $14,000 \times g$ for 10 min at 4°C. The supernatant was discarded and the DNA pellet was resuspended in TE buffer. DNA libraries were prepared using the Oxford Nanopore Ligation Sequencing Kit (SQK-LSK110) and library loading and sequencing were performed according to the manufacturer's instructions and protocol for Flongle.



2.2 Genome sequencing and assembly pipeline

For MinION data in use an in-house pipeline with tools suggested by Oxford Nanopore. The acquisition of the raw data, a series of processing steps were followed until these genomes were assembled. The first stage of data processing consisted of converting the raw signals into DNA base sequences (base calling) using the Guppy tool (Wick et al., 2019) and Dorado with super high accuracy settings. With this data in hand, the quality of the readings obtained in the previous step was assessed using the NanoStat software (De Coster et al., 2018). With these results, the processing continued with the removal of adapters from the base called reads to prepare them for subsequent steps; this was done using the Porechop tool (<https://github.com/rrwick/Porechop>), which also compressed the resulting reads. Once the adapters had been removed, a new quality analysis was carried out using the NanoStat tool. Once this step was completed, the low-quality bases and short reads were trimmed using the Chopper tool (De Coster and Rademakers, 2023), which not only trimmed but also compressed the resulting reads. The next step was to analyze the quality of the resulting sequence, again using the NanoStat tool. With the resulting data, we then used the Flye (Kolmogorov et al., 2019) and Unicycler (Wick et al., 2017a) tools to assemble the genomes using filtered, high-quality reads. The genome assembly was then refined using information from reads mapped using the Racon tool (Vaser et al., 2017). Finally, with the final result in hand, the quality of the final genome assembly was assessed using the Quast tool (Gurevich et al., 2013) and BUSCO (Manni et al., 2021). For Illumina procedures, BRA006 genome samples were sequenced using MiSeq technology at Macrogen facility (Seoul, South Korea), all the processing steps such as read mapping, trimming low quality reads and *de novo* genome assembly were performed using the proprietary software Geneious (Version 11) (Kearse et al., 2012).

2.3 Phylogenetic analysis

For phylogenetic inference, we choose a comprehensive method called digital DNA:DNA hybridization (dDDH) available in the Type Genome Server (TYGS) web tool (Meier-Kolthoff and Göker, 2019). We used both MinION and Illumina data as query sequences with a standard parameter set.

2.4 BGCs analysis and functional annotation

To reveal biosynthetic gene clusters (BGC) from the BRA006 genome, we used the antiSMASH tool (version 7.1.0) (Blin et al., 2023) with relaxed detection strictness and all extra features selection. Coding sequences (CDS) prediction of BRA006 assemblies were made using Prokka (1.14.6) (Seemann, 2014), which is based on prodigal (Hyatt et al., 2010) HMM models to identify proteins by their family motifs. Finally, we combine Prokka and antiSMASH results to obtain a better resolution of protein functions within BGCs. We manually filtered antiSMASH “Most Similar Known Cluster” feature and retrieved BGCs that matched MiBiG clusters related to functions of interest (version 3.0) (Terlouw et al.,

2023). Thus, python scripts were used to convert CDS of GenBank files from MiBiG and antiSMASH into fasta format in which the MiBiG sequences were used to construct reference databases and antiSMASH's as query sequences for BLASTp. With the TSV files from BLASTp, we grouped all possible matches by each query sequence and selected the one with the lowest e-value. Finally, we used BioPython to plot the comparisons with more than 70% of similarity.

2.5 Metabolomics analysis

The BRA006 isolates were cultivated in 100 mL of the sterile A1 culture medium (Starch 10 g/L, Yeast extract 4 g/L, Peptone 2 g/L and Sea Water 75%), in 250 mL Erlenmeyer flasks. After 5–7 days under 120 rpm agitation and 28°C, the liquid cultures were extracted with ethyl acetate (1:1) under agitation for 2 h, and the organic phase was dried under pressure and kept at 4°C. For the LC-MS/MS analyses, the organic extracts were diluted in methanol, and the extract solution to be injected was prepared in methanol:water solution at a ratio of 1:1 (v/v) with a final concentration of 1.0 mg/mL (Bauermeister et al., 2016). The LC-MS/MS analysis itself was conducted in the Acquity UPLC H-Class (Waters, Milford, MA - US) hyphenated with Impact II mass spectrometer (Bruker Daltonics, Billerica - US). The mobile phase (flow 0.3 mL.min⁻¹) consisted of water (A) and methanol (B) in the following gradient: 0.0–15.0 min (5%–20% B, curve 6); 15.0–30.0 min (20%–95% B, curve 6); 30.0–33.0 min - (100 B, curve 1); 33.0–40.0 min (5% B, curve 1). C18 - Luna (Phenomenex® - 100 mm × 2.1 mm × 2.6 μm) and the temperature adjusted to 35°C. The parameters adjusted for the spectrometer were: end plate offset of 500V; capillary voltage of 4.5kV; nitrogen (N₂) was used as gas; drying gas flow at 5.0 L.min⁻¹; drying gas temperature at 180°C; 4 bar nebulizer gas pressure; positive ESI mode. Spectra (*m/z* 30–2000) were recorded at a rate of 8 Hz. The quadrupole ion energy was set to 5.0 eV. The collision cell was set to 5.0 eV, with collision energies ranging from 20 to 50 eV (transfer time from 20 to 70 μs), and absolute fragmentation cutoff of 1,000 counts. Ions below 200 Da were excluded from fragmentation. The “active exclusion” function was enabled and with the following settings: exclude after 3 spectra; release after 0.3 min; reconsider precursor if the ratio current intensity to previous intensity was 1.8 (Bazzano et al., 2024). Accurate masses were obtained using sodium formate solution (10 mM) as a calibrant (Hoang et al., 2024).

2.6 GNPS2 molecular networking and *in silico* annotation

The .csv and .mgf files generated in MZmine3 (Schmid et al., 2023) from the raw data of the metabolomics analysis were imported into the GNPS2 platform (<https://gnps2.org/>), where the molecular network and spectral pairing were performed using the Feature Based Molecular Network (FBMN) (Nothias et al., 2020) (https://gnps2.org/workflowinput?workflowname=feature_based_molecular_networking_workflow). We used the standard parameters for FBMN. Once the network was built, the annotations were propagated using the ChemWalker (Borelli et al., 2023) tool through GNPS2 interface (<https://gnps2.org/workflowinput?>

TABLE 1 Summary information of Prokka and Busco tools applied on BRA006 genome.

Annotated proteins	With COG ^a	Hypothetical proteins	tRNA	rRNA	ncRNA	BUSCO ^b			Technique
						C	F	M	
11072	1,529	8,318	65	6	1	201	93	62	MinION
6,080	1769	3,289	69	5	1	350	1	5	Illumina
6,474	1807	3,601	68	9	1	328	17	11	MinION - corrected

This table contains the summary information of Prokka and Busco tools from the assembly of Illumina and MiniON, data sequencing techniques. The name "MinION, corrected" refers to the results of a troubleshooting process by Dorado and Unicycler, respectively. To remove base calling errors from the original MiniON, sequencing data.

^aComplete distributions of COG (Cluster of Orthologous Groups) functional categories for each assembly technique.

^bAssembly completeness analysis based on near-universal single-copy orthologs gene content by BUSCO. The letters C, F, and M stand for complete, fragmented, and missing sequences, respectively.

workflowname=chemwalker_nextflow_workflow). We used the standard parameters for ChemWalker, including COCONUT (Sorokina et al., 2021) as the reference database and 0 for the component index to propagate information for the whole network. For the nodes that could not be annotated from these two previous methods, the MS/MS mass spectra of those compounds were analyzed using the SIRIUS tool version 5.8.5. To use this tool, we followed the annotation recommendations for Q-TOF mass analyzers and adopt the following adducts $[M + H]^+$, $[M + K]^+$, and $[M + Na]^+$ as well as all the formulas available through the tool access (Dührkop et al., 2019). For *in silico* annotation, both ChemWalker and SIRIUS were used to annotate a structure with the best ranked candidate.

3 Results

3.1 Genomic analysis

The isolate BRA006 exhibits very characteristic growth. The colonies are orange in color with apical growth and a rough appearance with the presence of individual dark-colored spores. In solid medium, it releases as yet undetermined compounds that diffuse into the agar, giving it a pink-purple color, as can be seen in Supplementary Figure 1.

We compared Illumina and MinION whole genome sequence techniques results from BRA006. While Illumina assembly yielded 10 contigs and 6,734,372 base pairs (6.73 Mb) in total, sequencing by MinION showed higher genome completeness, with an assembly resulting in 4 contigs with 6,762,267 base pairs (6.76 Mb) in total. Functional annotation performed by Prokka showed a significant increase in CDS predicted from MinION assembly data driven by a high amount of hypothetical proteins. We corrected this issue adding modifications in MinION assembly pipeline and obtained similar values to Illumina. The complete result is shown in (Table 1). The CDS prediction followed the number of hypothetical proteins, with more predictions from MinION than from Illumina.

3.2 Metabolic potential

Micromonospora is one of the largest genera of Actinomycetota and possesses a large repertoire of bioactive secondary metabolites (SM) with a broad spectrum of therapeutic effects (Yan et al., 2022), for instance, aminoglycosides and macrolactam antibiotics. Through antiSMASH, we annotated the BGC content from Illumina and MinION assemblies of BRA006. MinION assembly resulted in a total of 15 BGCs (Table 2) that vary in similarity with antiSMASH database. Among them, there are those with reported antimicrobial, antifungal, and antitumor activities. For instance, quinolidomycin A is a macrolide with antibiotic and anticancer effects isolated from *Micromonospora* sp JY16 (Ruzzini and Clardy, 2016) and BRA006 presents a quinolidomycin BGC with 219 Mb in length and 67% similarity with the most similar known cluster.

The type-III polyketide Loseolamycin was identified from *Micromonospora* endolithica and inhibited the growth of the Gram-positive *Bacillus subtilis* and also showed herbicidal activity against the weed *Agrostis stolonifera* (Lasch et al., 2020). Cinerubins are anthracycline antimicrobials produced by actinomycetota which also present antitumor activity (Paderog et al., 2020; Silva et al., 2020). BRA006 possesses highly similar clusters to cinerubin B (74% by MinION and 80% by Illumina) and loseolamycin (88% by MinION and 92% by Illumina) A1/A2. BRA006 also has less similar BGCs for the biosynthesis of other natural compounds with antitumoral activity such as bleomycins and kedarcidin, isolated from *Streptomyces verticillus* and an unclassified Actinomycetales strain (ATCC 53650) Actinomycetes, respectively (Hecht, 2000; Hofstead et al., 1992).

AntiSMASH results from Illumina assembly yielded 18 BGCs (Table 2) in total, including clusters for the production of quinolidomycin, cinerubin B, and loseolamycin A1/A2 found in both Illumina and MinION data. However, only by sequencing with Illumina, it was possible to find a BGC 100% similar to the production of sungeidines (Low et al., 2020), a group of metabolites produced by pathways with close evolutionary relations with the antitumor dynemicins (Unno et al., 1997).

TABLE 2 Biosynthetic Gene Cluster from antiSMASH.

Region	Type	From	To	Most similar known cluster	Type from most similar known cluster	Similarity	Technique
1.1	T1PKS	173,653	262,468	catenulisporolides	NRP + Polyketide	12%	Illumina
1.2	T3PKS	972,718	1,013,770	loseolamycin A1/loseolamycin A2	Polyketide	92%/88% ^a	Both
1.3	thioamide-NRP	1,165,443	1,214,791	cadaside A/cadaside B	NRP	19%	Illumina
1.4	terpene	1,632,755	1,652,207	isorenieratene	Terpene	25%/25% ^a	Both
1.5	terpene	1,735,876	1,756,185	phosphonoglycans	Saccharide	3%	Illumina
1.9	T1PKS	3,818,978	3,889,418	quinolidomicin A	Polyketide	45%/67% ^a	Both
1.10	NI-siderophore	3,941,102	3,951,331	FW0622	Other	50%	Illumina
1.11	NRPS-like, NRPS, T1PKS, PKS-like	3,956,633	4,058,605	sungeidine C/sungeidine B/sungeidine D/sungeidine H/sungeidine A/sungeidine E/sungeidine F/sungeidine G	Polyketide	100%	Illumina
1.12	NRPS-like, NRPS, T1PKS	4,153,951	4,219,351	crochelin A	NRP + Polyketide	12%	Illumina
1.13	terpene	4,832,538	4,853,269	nocathiacin	RiPP:Thiopeptide	4%/4% ^a	Both
1.14	T2PKS	5,027,025	5,098,322	formicamycins A-M	Polyketide	18%	Illumina
1.15	oligosaccharide, terpene	5,268,386	5,304,696	lobosamide A/lobosamide B/lobosamide C	Polyketide	13%	Illumina
1.16	T2PKS, oligosaccharide, other, NRPS	5,309,290	5,436,713	cinerubin B	Polyketide:Type II polyketide	80%/74%	Both
1.17	NI-siderophore	5,616,661	5,629,872	peucechelin	NRP	10%	Illumina
1.18	terpene, RiPP-like	5,837,876	5,861,979	lymphostin/neolymphostinol B/lymphostinol/neolymphostin B	NRP + Polyketide	33%	Illumina
1.19	terpene	5,994,482	6,015,432	tetrachlorizine	Polyketide	13%	Illumina
1.20	other, ladderane, NRPS, arylpolyene	6,201,575	6,311,984	kedarcidin	NRP + Polyketide: Iterative type I polyketide + Polyketide:Enediyne type I polyketide	13%/6%	Both
6.1	NRPS-like, T1PKS	1	37,77	quinolidomicin A	Polyketide	28%/67%	Both

(Continued on the following page)

TABLE 2 (Continued) Biosynthetic Gene Cluster from antiSMASH.

Region	Type	From	To	Most similar known cluster	Type from most similar known cluster	Similarity	Technique
2.1	RiPP-like	136,664	147,482	lymphostin/ neolymphostinol B/lymphostinol/ neolymphostin b	Polyketide + NRP	15%	MinION
2.5	NI-siderophore	4,182,055	4,194,725	peucechelin	NRP	10%	MinION
3.2	NRPS-like	407,718	448,468	sarpeptin A/sarpeptin B	NRP	25%	MinION
3.4	oligosaccharide, terpene	530,94	565,187	brasilicardin A	Terpene + Saccharide	38%	MinION
3.5	T2PKS,NRPS- like	732,357	804,659	pradimicin-A	Polyketide	17%	MinION
3.7	NRPS	1,520,332	1,559,968	bleomycin A2/bleomycin B2	NRP + Polyketide + Saccharide	14%	MinION
3.9	T1PKS	1,799,831	1,841,948	rakicidin A/rakicidin B	NRP:Cyclic depsipeptide + Polyketide:Modular type I polyketide	40%	MinION
3.10	NI-siderophore	1,859,147	1,870,220	putrebactin/ avaroferrin	Other	50%	MinION

^aSimilarity data from Illumina and MinION, respectively.

3.3 Evolutionary relationships

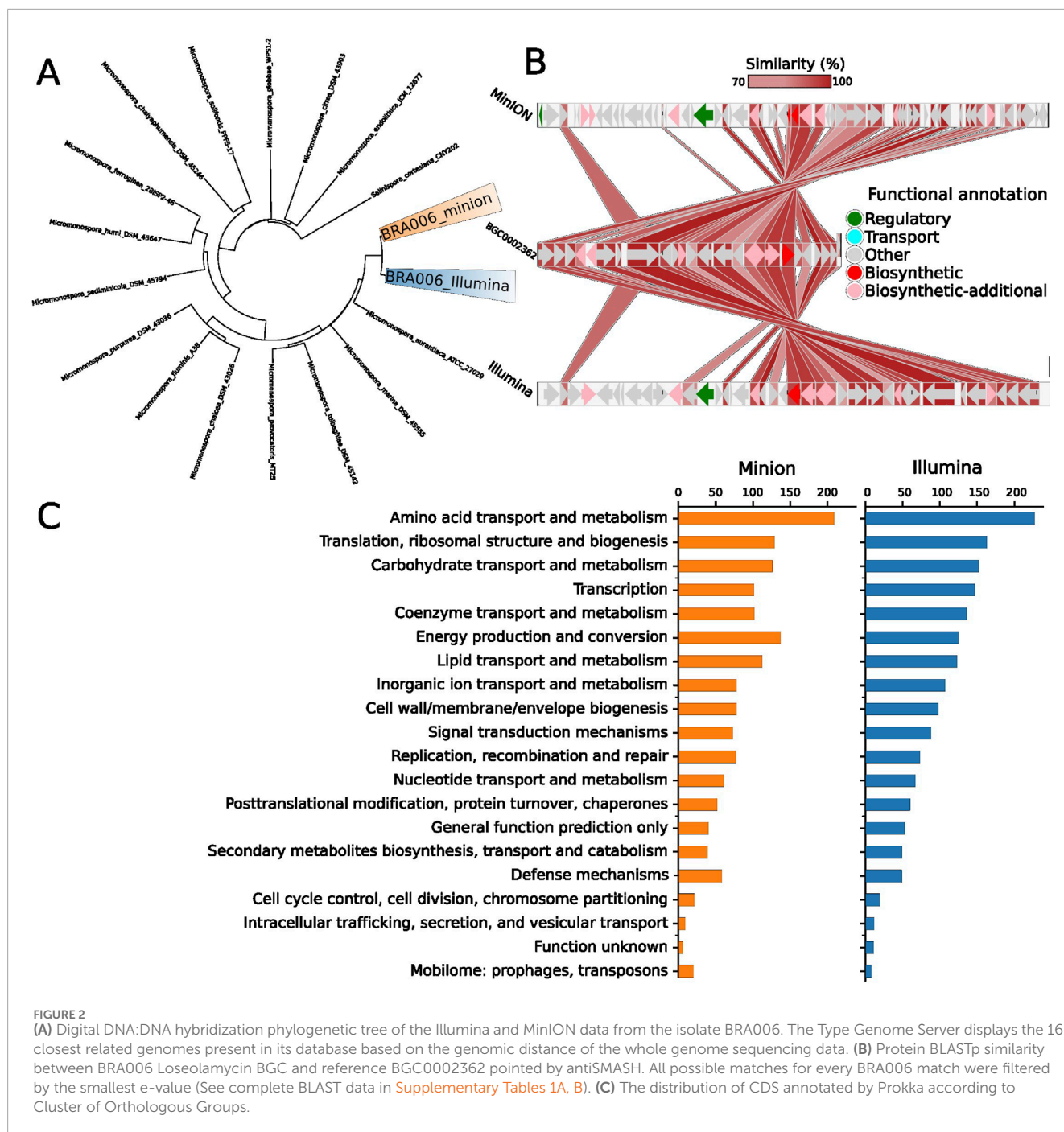
Since both genome sequencing methods yielded identical clusters with enzymes from pathways for synthesizing compounds with medical applications, we decided to explore the evolutionary relationship. According to digital DNA:DNA hybridization used by Type Genome Server (TYGS) (Meier-Kolthoff and Göker, 2019) for phylogenetic inferencing, these two assemblies were classified as *Micromonospora sp* and pointed out as possible novel species due to their relatively high genomic distance to its closest related group: *Micromonospora aurantiaca* ATCC 27029 (Figure 2). However, the BGC with higher similarity to the antiSMASH database is the one for the production of loseolamycin A1/A2 in both MinION and Illumina assemblies. Therefore we used BLASTp to compare the proteins within those BGCs to the reference: BGC0002362 from *Micromonospora endolithica*.

Figure 2B shows the protein similarity between loseolamycin-producing BGC from BRA006 and *M. endolithica*. Both assembly methods detected the complete inversion of this *M. endolithica* BGC followed by a series of indels in the upstream region majorly composed of CDSs that encode proteins with no functional annotation. Other BGCs with less similarity to the antiSMASH database were also compared with the reference sequences. For the cinebubin B and quinalidomicin A BGCs there were differences between antiSMASH results regarding the sequencing method. In the case of cinerubin B BGC, sequencing by Illumina resulted in a BGC with 172 Mb length

while MinION's with 71 Mb. Both with a few highly similar proteins to the reference BGS (BGC0000212) (Supplementary Figure 2) original from *Streptomyces sp.* SPBO074. On the other hand, quinolidomicin A BGCs showed 219 Mb in MinION and 70 Mb in Illumina assembly. Quinolidomicin A BGC from MinION data presented more similarity (67% against 45% from the counterpart) to reference BGC0002520 original from *Micromonospora sp* (Supplementary Figure 3) although with shorter CDSs.

3.4 Metabolomic analysis

From the analysis of potential BGCs found in BRA006, where their potential to biosynthesize compounds with antibacterial (loseolamycins A, quinolidomicin A, cinerubin B, and brasilicardin A), antifungal (pradimicin A) and anticancer (bleomycin A2 and kedarcidin) activity was identified, we investigated whether these compounds were present in the metabolome of this Actinomycetota. To do so, and to extend the analysis to other compounds that BRA006 might be able to produce, we performed a metabolomics analysis in which the BRA006 extract was analyzed by LC-ESI(+)-HRMS/MS. After the LC-MS/MS analysis, the obtained raw data were converted into. mzXML format using Proteowizard (Chambers et al., 2012), and feature finding was performed using MZmine (Schmid et al., 2023). We then performed the annotations sequentially, in three steps. In the first, the annotations were made



by spectral pairing and molecular network construction through GNPS2 (<https://gnps2.org/>). As a result, we obtained 527 nodes with at least one connection, of which 83 were annotated by GNPS2. Based on this result, we propagated the annotations to the nodes that did not present an annotation by spectral pairing using ChemWalker (Borelli et al., 2023), increasing the annotations to 373 nodes. Finally, for the nodes that could not be annotated by this tool, we used a third *in silico* spectral annotation tool, SIRIUS (Dührkop et al., 2019), increasing the annotations for an additional 67 nodes. Thus, a total of 527 nodes were represented by the molecular network, of which 523 were annotated. With these results in hand, we performed

an automated chemical classification of each annotated compound using ClassyFire (Djoumbou Feunang et al., 2016). The molecular network was colored based on the superclasses to which each annotated compound belonged (Figure 3).

The compounds could be grouped into 13 different chemical superclasses, where 30.2% of the annotated compounds belong to the superclass of lipids and lipid-like molecules, 19.5% to organoheterocyclic compounds, 15.1% to organic acids and derivatives, 9.9% to benzenoids, 7.6% to phenylpropanoids and polyketides, 6.5% to organic oxygen compounds, 4.0% to organic nitrogen compounds, 1.7% to alkaloids and derivatives, 0.6% to

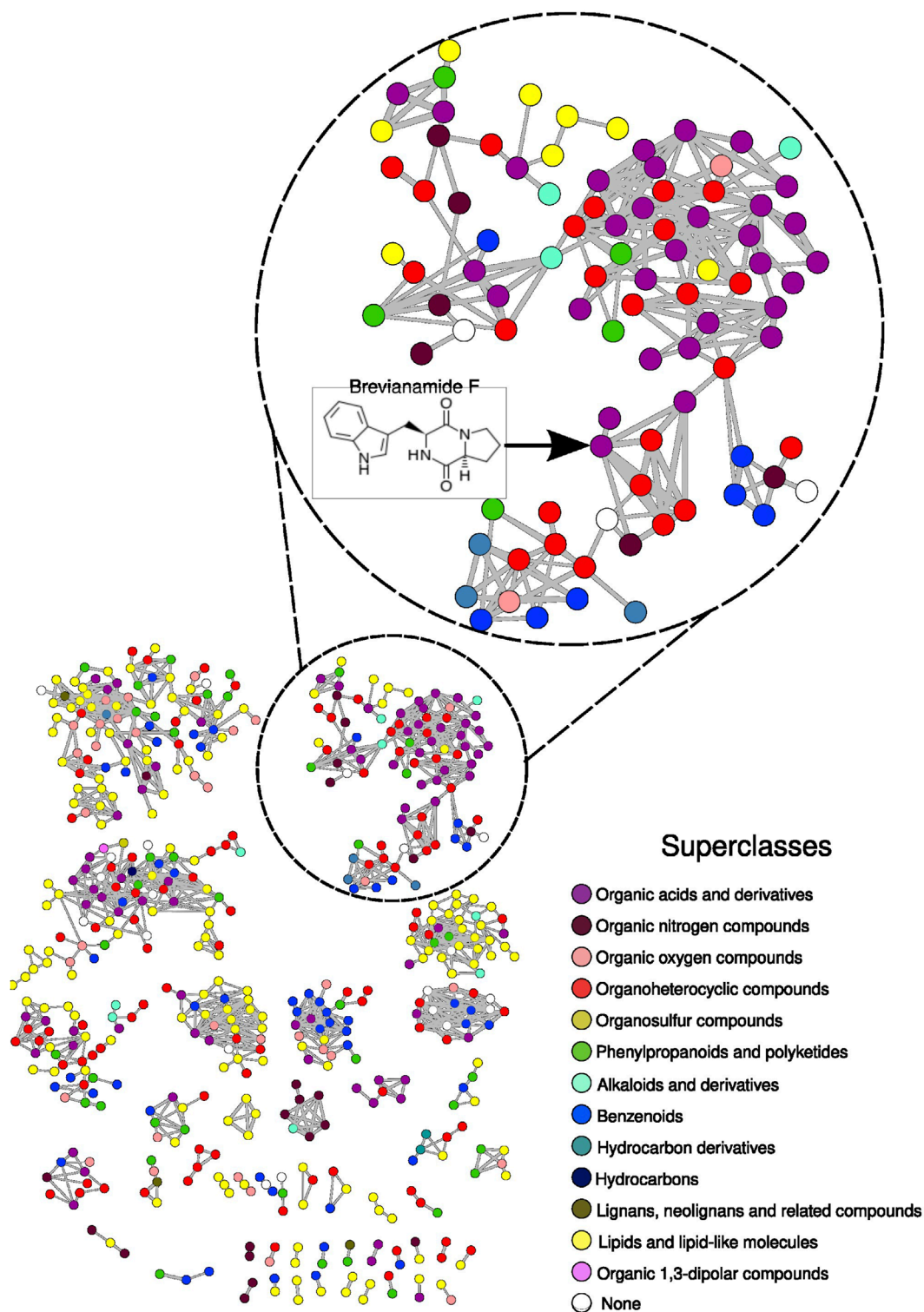


FIGURE 3

Molecular network constructed from BRA006 metabolomic data. Nodes were colored based on the superclass classification performed on the molecules annotated by ClassyFire. The full description of the annotation set is presented in [Supplementary Table 4](#). The cluster showing the node with the annotation for brevianamide F is highlighted.

lignans, neolignans, and related compounds, 0.4% to hydrocarbon derivatives, 0.2% to organic 1, 3-dipolar compounds, 0.2% to hydrocarbons, 0.2% to organosulfur compounds, and 3.8% could

not be classified (None). The complete relationship between all annotated compounds and their chemical hierarchical classification into kingdom, superclass, and class is shown in [Supplementary Table 4](#).

From the entire set of annotations and classifications, we searched for the seven compounds predicted by antiSMASH in the BRA006 metabolome. None of the seven compounds were found in the annotation pool. Therefore, we took the molecular structure of these seven compounds and performed their chemical classification using ClassyFire; once we had their chemical class, we would search for compounds annotated in the BRA006 metabolome that had the same chemical class. From a pharmacological point of view, compounds belonging to the same chemical class might belong to the same biosynthetic pathway and/or have similar effects, as is the case, for example, with the class of peptidomimetics in the treatment of cancer (de Valk et al., 2020) and steroids in the treatment of pain (Paulsen et al., 2013). The seven compounds were organized into six different classes of molecules, where loseolamycin A1 belongs to the phenol class, cinerubin B belongs to the anthracycline class, brasilicardin A belongs to the steroid and steroid derivative class, pradimicin A belongs to the naphthacene class, bleomycin A2 belongs to the peptidomimetic class, and quinolidomycin A and kedarcidin are organooxygen compounds.

Of these six chemical classes to which the compounds of interest belong, the Phenols class presented 10 annotated molecules, while the Organooxygen Compounds and Steroids and Steroid Derivatives classes presented 34 compounds each and Peptidomimetics presented two compounds in our analysis. Among the 10 compounds belonging to the Phenol class, three of them had bioactivity previously reported, but with a different action from the antibiotic loseolamycins A1. The targeting offered by antiSMASH, by searching for molecules belonging to the same class as those with bioactive activity predicted by the tool, allowed us to find a larger and more diverse range of compounds.

Using a reverse flow of integrative analysis, where we start from what was annotated in the metabolome, we set out to evaluate whether it would be possible to identify, from a given metabolite, the enzymes that lead to its production in BRA006. To this end, we first crossed the annotated metabolome with the KEGG database (Kanehisa and Goto, 2000) to search within the metabolome of this Actinomycetota for molecules with known biosynthetic pathways. As a result, we found a molecule belonging to the staurosporin biosynthetic pathway, brevianamide F. It should be noted that of the three methods used in the annotation process, brevianamide F was annotated by spectral pairing with the GNPS spectral library (Wang et al., 2016), showing a MQScore of 0.97 and a m/z error of 2.47 ppm. Since the topology of the molecular network is given by the similarity between nodes, the node with an annotation for brevianamide F is connected to seven other nodes (Figure 3). All seven nodes have been annotated by ChemWalker. Looking at the predicted structures, six of the seven share the same indole-like nucleus as brevianamide F, suggesting that other molecules may ultimately be produced, either by brevianamide F BGC or by other intermediates belonging to the same pathway as brevianamide F. Once identified, the enzymes that make up the staurosporin biosynthetic pathway, we searched the BRA006 genome for which of these enzymes would be encoded. From this search, 16 staurosporin pathway enzymes were identified in the genome of this Actinomycetota (Supplementary Figure 4), and 11 overlap the region of the BGC 20 from Illumina (Figure 4). Among them, we found an NRPK 2,3-dihydroxybenzoate-AMP ligase functionally classified as a biosynthetic-additional enzyme by antiSMASH. These results show that a dynamic integrative

approach, i.e., first combining spectral and *in silico* annotation, assigning chemical classes, and then searching for metabolites from the genome to match encoded proteins in the genome to the pathway producing the putative metabolite annotated, is an efficient approach to characterize new species with the potential to produce bioactive compounds.

4 Discussion

The multiomics characterization of new *Micromonospora* strains, especially those found in the marine environment, is a highly relevant task given the potential for the discovery of new bioactive compounds. Multiomics approaches allow combining information from different sources and amplify the scope of features that can be learnt (Kwoji et al., 2023). For instance, genomic annotation are hypothesis on the proteins that a living organism can produce, although there is no information about the levels nor the conditions in which these proteins are expressed. On the other hand, metabolomics inform the set of compounds conditionally produced but without information on the enzymes or metabolic pathways (O'Donnell et al., 2020). Metabologenomics, integrates genomic and metabolomics and relate the compounds found to the enzymes encoded in the genome (Go et al., 2024).

In this sense, the use of dynamic integrative analytical approaches seems to be a promising resource, both in the process of characterizing new species and in the evaluation of the potential for the production of natural products from a new microorganism.

In the present work, we characterized BRA006, a potential new species of the genus *Micromonospora* in terms of secondary metabolites production. We used both genomic and metabolomic points of view, comparing two whole genome sequencing approaches (Figure 2A) and a multiple-step metabolite annotation workflow. This strategy introduces an innovative dynamic approach to multiomics analysis in which the annotated BGCs types predicted by antiSMASH guided the search for bioactive compounds in the metabolome content, as well as the compounds identified in the metabolome (which match specific pathways) led to the search for CDSs that encode enzymes from these pathways. Among the various microbial genera described to date that stand out for their ability to produce bioactive compounds, the genus *Micromonospora* is an important model in natural products research and a milestone in the discovery of new biocompounds (Hifnawy et al., 2020). The potential to produce bioactive natural products from bacterial isolates from the Brazilian coast is already known, especially those with antitumor activity (Sousa et al., 2012; Silva et al., 2017). Among these, the activity observed in crude extracts of the genus *Micromonospora* was attributed to a group of anthracyclonones (Sousa et al., 2012).

According to Yan et al. (2022), species from the *Micromonospora* genus encode from 4,200 to 8,017 proteins and have genome sizes from 5.07 to 9.24 Mb. BRA006 possesses a 6.7 Mb genome, confirmed by two independent sequencing methods, although MinION assembly annotated by Prokka showed up to 11,000 CDS against 6,080 from Illumina. This gap between them is mostly due to higher error rates from MinION assembly, which probably causes artificial stop codons that could explain the higher number of CDSs. We decreased this gap changing our basecaller and

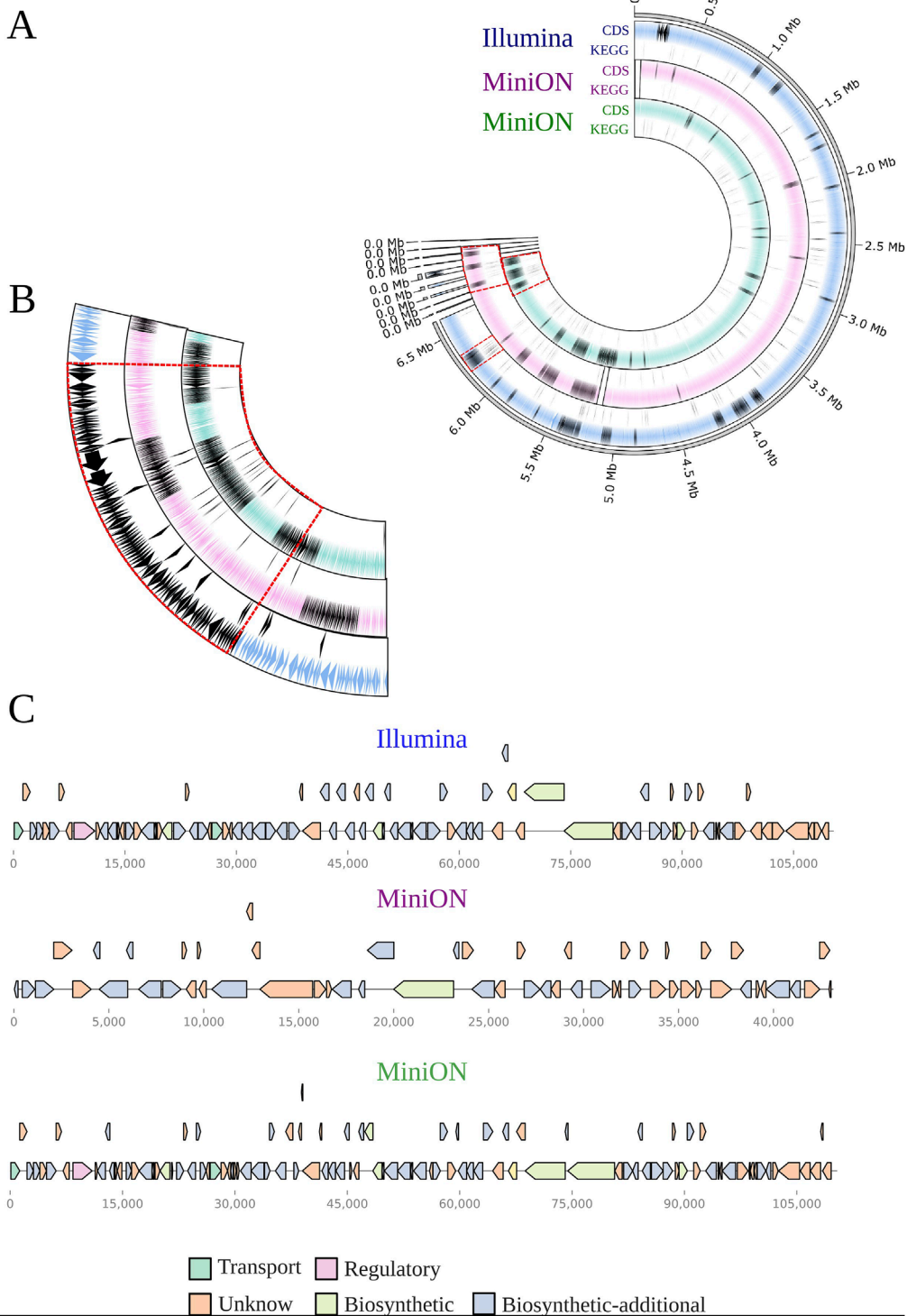


FIGURE 4
Circular representation of BRA006 genome according to the sequencing technique. **(A)** Each circle contains a track with the CDS predicted by prokka with BGCs annotated by AntiSMASH and a underneath track highlighting the proteins present in the KO00404 map from KEGG (<https://www.genome.jp/pathway/ko00404>). The pink-colored track contains results from MinION assembly pipeline with Guppy and Fly, whilst the green tracks refers to pipeline with Dorado and Unicycler. **(B)** Comparison among BGC. We set BGC number 20 from Illumina data as que reference to show the fragmented BGCs from MinION data. **(C)** Detailed visualization of Illumina BGC 20 and their respective corresponding from MinION.

genome assembler tools from Gumpy to Dorado and from Flyer to Unicycler, respectively. However, this still resulted in a fragmented kedarcidin-producing BGC which overlaps with the genes that encode enzymes that are also part of the staurosporin pathway. An alternative to solve this issue would be to perform a hybrid assembly (Wick et al., 2017b; Laver et al., 2015). Besides the difference in CDS number, antiSMASH found the same BGCs in both assemblies, such as Cinerubin B and Loseolamycins, with high similarity to antiSMASH database. Also, the whole genome sequence phylogeny placed both assemblies as a monophyletic group dissimilar enough from *Micromonospora aurantiaca* ATCC27029 to TYGS to point BRA006 as a possible novel *Micromonospora* species. As an example of genetic divergence of BRA006 from other *Micromonospora*, we can cite the loseolamycin A1 BGC, where it is possible to observe a complete inversion of the cluster and several indels. According to Medema et al. (2014) NRPK clusters evolved from gene duplication followed by differentiation, which could explain the difference between BRA006 and *M. endolithica*. Unfortunately, even with Prokka, annotating most of the proteins in that BGC was not yet possible.

The approach “from genomics to metabolomics” yielded a BGC that encodes pathways to compounds with pharmaceutical applications, which confirms the importance of the *Micromonospora* genus, although the compounds produced by these BGCs were not found in metabolomic data. Their absence can be explained by differences between laboratory culture media and the original ecological niche, as well as the need for improvements in the acquisition parameters. Genomic-guided works often require heterologous expression of parts or the entire BGC to obtain the active compound in the laboratory (Xu and Wright, 2019). For instance, Lasch et al. (2020) obtained loseolamycins from *M. endolithica* by heterologously expressing type III polyketide synthase, Domingues Vieira et al. increased the production of eponemycin and related epoxyketone peptides by cloning the whole epn-tmc BGC from *Streptomyces* sp. BRA346 (Domingues Vieira et al., 2022), and Yamanaka produced Taraomycin A by editing regulators of this BGC from *Saccharomonospora* sp. CNQ490 (Yamanaka et al., 2014).

Starting from metabolomics, the analysis of the BRA006 metabolome allowed us to identify annotated compounds belonging to already well-established chemical classes whose biosynthesis is reported in the literature for this genus, as in the case of macrolides (Hifnawy et al., 2020). Of these, we were able to annotate nine compounds belonging to this class (Supplementary Table 4), two of which previously reported bioactivity: tricholide A, with antibacterial activity (Bertin et al., 2017) and 11,12-dihydroxy-6,14-dimethyl-1,7-dioxacyclotetradeca-3,9-diene-2,8-dione, with immunosuppressive activity (Fujimoto et al., 1998). It should be noted that two compounds were recorded as tricholide A. Both presented the same *m/z* value but with very different retention times, indicating the presence of isomers of this compound, as both appear as neighboring nodes in the molecular network, and also which can be seen by the extratec ion chromatogram (XIC) (Supplementary Figure 5B). The annotation procedures and KEGG pathway search identified Brevianamide F, a compound with activity against *Staphylococcus aureus* (Ben Ameer Mehdi et al., 2009) and an intermediate of well-established inducer of apoptosis, Staurosporin (Belmokhtar et al., 2001). Among the three annotation

methods for a given molecule we used, spectral matching against a reference library is the best available resource (Ausloos et al., 1999). In addition, we used two other *in silico* annotation resources: ChemWalker and SIRIUS. Of these two tools, SIRIUS has the best accuracy, but it is very difficult to use when dealing with large sets of spectra. This limitation is overcome by ChemWalker, which allows greater annotation coverage, taking into account the topology of the molecular network. We reached brevianamide F annotation through two different annotation routes, which brings robustness to the interpretation of the result obtained and highlights the potential for a reverse flow in elucidating the biosynthetic potential of a new non-model organism. The spectral library match was reached through pairing spectra on GNPS2, where the results on XIC as well as the mirrorplot for this annotated compound is presented in Supplementary Figure 5. It is worth highlighting that we also carried out an *in silico* prediction analysis using SIRIUS for the seven nodes related to that of Brevianimide F. However, the best candidates predicted by this tool had little structural similarity with the compound itself, unlike the candidates provided by ChemWalker. The advantage of ChemWalker is that it uses the sample context given by the molecular network topology on compound re-ranking. Eventually, this new proposed BGC could synthesize the other molecules annotated and linked to the brevianamide F's node or, maybe, they are intermediates from different pathways that had not been described yet. In both cases, the knowledge of these analogues opens the possibilities to improve the known bioactivity of Brevianamide F, which could be tested by isolation or even by (bio)synthesizing the compounds.

Inspecting KEGG's metabolic pathways, we found that brevianamide F, a product of fungi metabolism (Mehetre et al., 2019), is part of staurosporin biosynthesis (KO00404). Therefore, we retrieved all EC numbers from the KO00404 pathway, connected them to our Prokka data, and found three enzymes that can catalyze the Brevianamide F synthesis reaction. In the KO00404 pathway, brevianamide F biosynthesis requires tryptophan and proline as substrates, being its core assembled by a non-ribosomal peptide synthetase (COG1020), named brevianamide F synthase (EC: 6.3.2.-), which is encoded by the gene *FtmA* (NCBI ID *Aspergillus fumigatus* (AFUA_8G00170) (Wang et al., 2023), and is also reported in *Streptomyces* sp (Maiya et al., 2006). The isolate BRA006 has an NRPK *mbtB_1* (MinION data) that matches with COG and EC number of *FtmA*, but is not a component of any BGC found by antiSMASH. However, examining the downstream and upstream regions of 2,3-dihydroxybenzoate-AMP ligase CDS (Illumina data) we found a genomic region that has the potential to be part of the brevianamide F synthesis pathway BGC.

AntiSMASH identifies BGCs based on profiles of Hidden Markov Models (pHMM) from PFAM (Mistry et al., 2021), TIGRFAMS (Haft et al., 2013), SMART (Letunic et al., 2021), BAGEL (van Heel et al., 2018; Yadav et al., 2009) and custom models that recognize signature sequences of such conserved domains in genomic query sequences (Biermann et al., 2022). However, there are BGCs that lack universal class-specific signature sequences and therefore are partially identified. To overcome this limitation, deep-learning-based tools such as DeepBGC (Hannigan et al., 2019) have been applied in genomic mining research to uncover new BGCs.

It is interesting to emphasize the innovative approach used in the present work with a two-way analysis of the genome

and metabolome. We started with the metabolome to see if the biosynthetic gene clusters involved in the production of a given metabolite could be identified in the genome. We then analyzed the genome sequencing data, and from there, by searching specific databases such as antiSMASH, we went to the metabolome to check whether the compounds predicted by antiSMASH were being produced (Domingues Vieira et al., 2022). Traditionally, the search for potential new compounds with bioactivity follows the latter linear flow of analysis, which in our case did not result in the identification of 7 of the metabolites predicted by genome mining. However, by using the chemical classes of these compounds, we could find analogues in our metabolomic data.

In addition, we present a new approach that integrates the classical approach with a reverse analysis, starting from the metabolome to the genome. For example, in neither short-read (Illumina) nor long-read (minION) sequencing data, it was possible to automatically detect the biosynthetic gene cluster for brevianamide F or staurosporin production. By using the two-way approach we could identify brevianamide F (reported as an intermediate in the Staurosporin biosynthesis) in BRA006 metabolome and, from there, we could identify some of brevianamide F' putative analogs and a BGC in the BRA006 genome, initially annotated with other function, that could represent Brevianamide F biosynthetic pathway in BRA006. Therefore, the approach presented in the present work allowed us to extend the characterization of the potential of bioactive natural products produced by BRA006.

Data availability statement

All mass spectrometry metabolomics data are available at MassIVE through the identifier: MSV000095044. The parameters used for preprocessing in MZmine3 are available at Zenodo <https://doi.org/10.5281/zenodo.10366840>. All data, Python scripts and Jupyter notebooks used during multiomics data analysis are available on this project GitHub page: <https://github.com/computational-chemical-biology/metabologenomics>. The FBMN results can be found here <https://gnps2.org/status?task=7b134da60f0f4a80aec790d2a294aedd> and ChemWalker results here <https://gnps2.org/status?task=9141a5cdabf842d39387e514e5305398>. The assemblies are available on NCBI database. MiION: <https://www.ncbi.nlm.nih.gov/biosample/SAMN39609461/> and Illumina: <https://www.ncbi.nlm.nih.gov/biosample/?term=SAMN29586427>.

Author contributions

GA: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing–original draft, Writing–review and editing. TB: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing–original draft, Writing–review and editing. EF: Data curation, Investigation, Validation, Writing–original draft, Writing–review and editing. RdF: Data curation, Investigation, Methodology, Validation,

Writing–original draft, Writing–review and editing. PR-T: Data curation, Writing–original draft, Writing–review and editing. MP: Data curation, Writing–original draft, Writing–review and editing. FR: Data curation, Writing–original draft, Writing–review and editing. GV: Data curation, Writing–original draft, Writing–review and editing. LM: Data curation, Formal Analysis, Software, Visualization, Writing–original draft, Writing–review and editing. HT: Data curation, Software, Writing–original draft, Writing–review and editing. LN: Data curation, Writing–original draft, Writing–review and editing. LG: Data curation, Writing–original draft, Writing–review and editing. GP: Data curation, Writing–original draft, Writing–review and editing. AG-d-l-F: Data curation, Software, Validation, Writing–original draft, Writing–review and editing. MW: Funding acquisition, Resources, Supervision, Writing–original draft, Writing–review and editing. NL: Funding acquisition, Project administration, Resources, Supervision, Writing–original draft, Writing–review and editing. DB: Funding acquisition, Project administration, Resources, Writing–original draft, Writing–review and editing. LC-L: Funding acquisition, Project administration, Resources, Writing–original draft, Writing–review and editing. M-EG: Writing–review and editing, Funding acquisition, Project administration, Resources, Writing–original draft. RR: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing–original draft, Writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The authors declare that they have received financial support from the following agencies São Paulo State Foundation (FAPESP, Award 2017/18922-2; 2020/02207-5; 2021/08235-3; 2021/10401-9; 2021/01748-5; 2021/09375-3; 2019/25432-7). M-EG was supported by CNPq Research Productivity Scholarship (award 302750/2020-7). MW was supported by the UCR startup. GA was supported by CNPq (141263/2021-0).

Acknowledgments

Sample collection and bioprospection authorizations were granted by the Biodiversity Authorization and Information System (SISBIO), number 26286-1.

Conflict of interest

MW is a co-founder of Ometa Labs LLC.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of *Frontiers*, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations,

References

- Andrade, L. S. N. (2020). *Identification and characterization of biosynthetic clusters from Micromonospora sp.* São Paulo, Brazil: M.Sc dissertation, University of São Paulo.
- Ausloos, P., Clifton, C. L., Lias, S. G., Mikaya, A. I., Stein, S. E., Tchekhovskoi, D. V., et al. (1999). The critical evaluation of a comprehensive mass spectral library. *J. Am. Soc. Mass Spectrom.* 10 (4), 287–299. doi:10.1016/S1044-0305(98)00159-7
- Bauermeister, A., Zucchi, T. D., and Moraes, L. A. (2016). Mass spectrometric approaches for the identification of anthracycline analogs produced by actinobacteria. *J. Mass Spectrom.* 51 (6), 437–445. doi:10.1002/jms.3772
- Bazzano, C. F., de Felício, R., Alves, L. F. G., Costa, J. H., Ortega, R., Vieira, B. D., et al. (2024). NP3 MS workflow: an open-source software system to empower natural product-based drug discovery using untargeted metabolomics. *Anal. Chem.* 96 (19), 7460–7469. doi:10.1021/acs.analchem.3c05829
- Behsaz, B., Bode, E., Gurevich, A., Shi, Y. N., Grundmann, F., Acharya, D., et al. (2021). Integrating genomics and metabolomics for scalable non-ribosomal peptide discovery. *Nat. Commun.* 12, 3225. doi:10.1038/s41467-021-23502-4
- Belmokhtar, C. A., Hillion, J., and Ségal-Bendirdjian, E. (2001). Staurosporine induces apoptosis through both caspase-dependent and caspase-independent mechanisms. *Oncogene* 26, 3354–3362. doi:10.1038/sj.onc.1204436
- Ben Ameer Mehdi, R., Shaaban, K. A., Rebai, I. K., Smaoui, S., Bejar, S., and Mellouli, L. (2009). Five naturally bioactive molecules including two rhamnopyranoside derivatives isolated from the Streptomyces sp. strain TN58. *Nat. Prod. Res.* 12, 1095–1107. doi:10.1080/14786410802362352
- Bertin, M. J., Roduit, A. F., Sun, J., Alves, G. E., Via, C. W., Gonzalez, M. A., et al. (2017). Tricholides A and B and unnarmicin D: new hybrid PKS-nrps macrocycles isolated from an environmental collection of trichodesmium thiebautii. *Mar. Drugs* 7, 206. doi:10.3390/md1507206
- Biermann, F., Wenski, S. L., and Helfrich, E. J. N. (2022). Navigating and expanding the roadmap of natural product genome mining tools. *Beilstein J. Org. Chem.* 18, 1656–1671. doi:10.3762/bjoc.18.178
- 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 Res.* 51, W46–W50. doi:10.1093/nar/gkad344
- Borelli, T. C., Arini, G. S., Feitosa, L. G. P., Dorrestein, P. C., Lopes, N. P., and da Silva, R. R. (2023). Improving annotation propagation on molecular networks through random walks: introducing ChemWalker. *Bioinformatics* 39 (3), btad078. doi:10.1093/bioinformatics/btad078
- Chambers, M. C., Maclean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., et al. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* 10, 918–920. doi:10.1038/nbt.2377
- De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., and Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 2018 (15), 2666–2669. doi:10.1093/bioinformatics/bty149
- De Coster, W., and Rademakers, R. (2023). NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* 5, btad311. doi:10.1093/bioinformatics/btad311
- de Valk, K. S., Deken, M. M., Handgraaf, H. J. M., Bhairasingh, S. S., Bijlstra, O. D., van Esdonk, M. J., et al. (2020). First-in-Human assessment of cRGD-zw800-1, a zwitterionic, integrin-targeted, near-infrared fluorescent peptide in colon carcinoma. *Clin. Cancer Res.* 15, 3990–3998. doi:10.1158/1078-0432.CCR-19-4156
- Djombou Feunang, Y., Eisner, R., Knox, C., Chepelev, L., Hastings, J., Owen, G., et al. (2016). ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *J. Cheminform* 8, 61. doi:10.1186/s13321-016-0174-y
- Domingues Vieira, B., Niero, H., de Felício, R., Giolo Alves, L. F., Freitas Bazzano, C., Sigrist, R., et al. (2022). Production of epoxyketone peptide-based proteasome inhibitors by Streptomyces sp. BRA-346: regulation and biosynthesis. *Front. Microbiol.* 13, 786008. doi:10.3389/fmicb.2022.786008
- Dührkop, K., Fleischauer, M., Ludwig, M., Aksenov, A. A., Melnik, A. V., Meusel, M., et al. (2019). SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Methods* 16 (4), 299–302. doi:10.1038/s41592-019-0344-8
- Finical, W., and Jensen, P. R. (2006). Developing a new resource for drug discovery: marine actinomycete bacteria. *Nat. Chem. Biol.* 12, 666–673. doi:10.1038/nchembio841
- Fujimoto, H., Nagano, J., Yamaguchi, K., and Yamazaki, M. (1998). Immunosuppressive components from an ascomycete, diplogelasinospora grovesii. *Chem. Pharm. Bull. (Tokyo)* 3, 423–429. doi:10.1248/cpb.46.423
- Go, D., Yeon, G. H., Park, S. J., Lee, Y., Koh, H. G., Koo, H., et al. (2024). Integration of metabolomics and other omics: from microbes to microbiome. *Appl. Microbiol. Biotechnol.* 108 (1), 538. doi:10.1007/s00253-024-13384-z
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 8, 1072–1075. doi:10.1093/bioinformatics/btt086
- Haft, D. H., Selengut, J. D., Richter, R. A., Harkins, D., Basu, M. K., and Beck, E. (2013). TIGRFAMs and genome properties in 2013. *Nucleic Acids Res.* 41, D387–D395. doi:10.1093/nar/gks1234
- Hannigan, G. D., Prihoda, D., Palicka, A., Soukup, J., Klempir, O., Rampula, L., et al. (2019). A deep learning genome-mining strategy for biosynthetic gene cluster prediction. *Nucleic Acids Res.* 18, e110. doi:10.1093/nar/gkz654
- Hecht, S. M. (2000). Bleomycin: new perspectives on the mechanism of action. *J. Nat. Prod.* 1, 158–168. doi:10.1021/np990549f
- Hifnawy, M. S., Fouda, M. M., Sayed, A. M., Mohammed, R., Hassan, H. M., AbouZid, S. E., et al. (2020). The genus Micromonospora as a model microorganism for bioactive natural product discovery. *RSC Adv.* 10, 20939–20959. doi:10.1039/d0ra04025h
- Hoang, C., Uritboonthai, W., Hoang, L., Billings, E. M., Aisporna, A., Nia, F. A., et al. (2024). Tandem mass spectrometry across platforms. *Anal. Chem.* 96 (14), 5478–5488. doi:10.1021/acs.analchem.3c05576
- Hofstead, S. J., Matson, J. A., Malacko, A. R., and Marquardt, H. (1992). Kedarcidin, a new chromoprotein antitumor antibiotic. II. Isolation, purification and physico-chemical properties. *J. Antibiot. (Tokyo)* 8, 1250–1254. doi:10.7164/antibiotics.45.1250
- Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinforma.* 1, 119. doi:10.1186/1471-2105-11-119
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28 (1), 27–30. doi:10.1093/nar/28.1.27
- Kato, N. N., Arini, G. S., Silva, R. R., Bichuette, M. E., Bitencourt, J. A. P., and Lopes, N. P. (2024). The world of cave microbiomes: biodiversity, ecological interactions, chemistry, and the multi-omics integration. *J. Braz. Chem. Soc.* 35, e-20230148. doi:10.21577/0103-5053.20230148
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 12, 1647–1649. doi:10.1093/bioinformatics/bts199
- Kim, L. J., Ohashi, M., Zhang, Z., Tan, D., Asay, M., Cascio, D., et al. (2021). Prospecting for natural products by genome mining and microcrystal electron diffraction. *Nat. Chem. Biol.* 17, 872–877. doi:10.1038/s41589-021-00834-2
- Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 5, 540–546. doi:10.1038/s41587-019-0072-8
- Kwoji, I. D., Aiyegoro, O. A., Okpeku, M., and Adeleke, M. A. (2023). 'Multi-omics' data integration: applications in probiotics studies. *NPJ Sci. Food* 5 (1), 25. doi:10.1038/s41538-023-00199-x
- Lasch, C., Gummerlich, N., Myronovskiy, M., Paluszczak, A., Zapp, J., and Luzhetskyy, A. (2020). Loseolamycins: a group of new bioactive alkylresorcinols produced after

or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2025.1515276/full#supplementary-material>

- heterologous expression of a type III PKS from *Micromonospora endolithica*. *Molecules* 20, 4594. doi:10.3390/molecules25204594
- Laver, T., Harrison, J., O'Neill, P. A., Moore, K., Farbos, A., Paszkiewicz, K., et al. (2015). Assessing the performance of the Oxford Nanopore technologies MinION. *Biomol. Detect. Quantif.* 3, 1–8. doi:10.1016/j.bdq.2015.02.001
- Letunic, I., Khedkar, S., and Bork, P. (2021). SMART: recent updates, new developments and status in 2020. *Nucleic Acids Res.* 49, D458–D460. doi:10.1093/nar/gkaa937
- Low, Z. J., Ma, G. L., Tran, H. T., Zou, Y., Xiong, J., Pang, L., et al. (2020). Sungeidines from a non-canonical enediyne biosynthetic pathway. *J. Am. Chem. Soc.* 4, 1673–1679. doi:10.1021/jacs.9b10086
- Maiya, S., Grundmann, A., Li, S. M., and Turner, G. (2006). The fumitremorgin gene cluster of *Aspergillus fumigatus*: identification of a gene encoding brevianamide F synthetase. *Chembiochem* 7, 1062–1069. doi:10.1002/cbic.200600003
- Manni, M., Berkeley, M. R., Seppey, M., and Zdobnov, E. M. (2021). BUSCO: assessing genomic data quality and beyond. *Curr. Protoc.* 12, e323. doi:10.1002/cpz1.323
- Medema, M. H., Cimermanic, P., Sali, A., Takano, E., and Fischbach, M. A. (2014). A systematic computational analysis of biosynthetic gene cluster evolution: lessons for engineering biosynthesis. *PLoS Comput. Biol.* 12, e1004016. doi:10.1371/journal.pcbi.1004016
- Mehetre, G. T. J. S. V., Burkul, B. B., Desai, D. B. S., Dharme, M. S., and Dastager, S. G. (2019). Bioactivities and molecular networking-based elucidation of metabolites of potent actinobacterial strains isolated from the Unkeshwar geothermal springs in India. *RSC Adv.* 17, 9850–9859. doi:10.1039/c8ra09449g
- Meier-Kolthoff, J. P., and Göker, M. (2019). TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat. Commun.* 1, 2182. doi:10.1038/s41467-019-10210-3
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2021). Pfam: the protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi:10.1093/nar/gkaa913
- Newman, D. J., and Cragg, G. M. (2020). Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* 83, 770–803. doi:10.1021/acs.jnatprod.9b01285
- Nothias, L. F., Petras, D., Schmid, R., Dührkop, K., Rainer, J., Sarvepalli, A., et al. (2020). Feature-based molecular networking in the GNPS analysis environment. *Nat. Methods* 17 (9), 905–908. doi:10.1038/s41592-020-0933-6
- O'Donnell, S. T., Ross, R. P., and Stanton, C. (2020). The progress of multi-omics technologies: determining function in lactic acid bacteria using a systems level approach. *Front. Microbiol.* 10, 3084. doi:10.3389/fmicb.2019.03084
- Paderog, M. J. V., Suarez, A. F. L., Sabido, E. M., Low, Z. J., Saludes, J. P., and Dalisay, D. S. (2020). Anthracycline shunt metabolites from philippine marine sediment-derived *Streptomyces* destroy cell membrane integrity of multidrug-resistant *Staphylococcus aureus*. *Front. Microbiol.* 11, 743. doi:10.3389/fmicb.2020.00743
- Paulsen, Ø., Aass, N., Kaasa, S., and Dale, O. (2013). Do corticosteroids provide analgesic effects in cancer patients? A systematic literature review. *J. Pain Symptom Manage* 1, 96–105. doi:10.1016/j.jpainsymman.2012.06.019
- Ruzzini, A. C., and Clardy, J. (2016). Gene flow and molecular innovation in bacteria. *Curr. Biol.* 26, R859–R864–R864. doi:10.1016/j.cub.2016.08.004
- Schmid, R., Heuckeroth, S., Korf, A., Smirnov, A., Myers, O., Dyrland, T. S., et al. (2023). Integrative analysis of multimodal mass spectrometry data in MZmine 3. *Nat. Biotechnol.* 41 (4), 447–449. doi:10.1038/s41587-023-01690-2
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 14, 2068–2069. doi:10.1093/bioinformatics/btu153
- Silva, A. E. T., Guimarães, L. A., Ferreira, E. G., Torres, M. da C. M., da Silva, A. B., Branco, P. C., et al. (2017). Bioprospecting anticancer compounds from the marine-derived actinobacteria *actinomadura* sp. collected at the saint peter and saint Paul archipelago (Brazil). *J. Braz. Chem. Soc.* 28, 465–474. doi:10.21577/0103-5053.20160297
- Silva, L. J., Crevelin, E. J., Souza, D. T., Lacerda-Júnior, G. V., de Oliveira, V. M., Ruiz, A. L. T. G., et al. (2020). Actinobacteria from Antarctica as a source for anticancer discovery. *Sci. Rep.* 1, 13870. doi:10.1038/s41598-020-69786-2
- Sorokina, M., Merseburger, P., Rajan, K., Yirik, M. A., and Steinbeck, C. (2021). COCONUT online: collection of open natural products database. *J. Cheminform* 13 (1), 2. doi:10.1186/s13321-020-00478-9
- Sousa, T. da S., Jimenez, P. C., Ferreira, E. G., Silveira, E. R., Braz-Filho, R., Pessoa, O. D., et al. (2012). Anthracyclines from *Micromonospora* sp. *J. Nat. Prod.* 75, 489–493. doi:10.1021/np200795p
- Sun, W., Wu, W., Liu, X., Zaleta-Pinet, D. A., and Clark, B. R. (2019). Bioactive compounds isolated from marine-derived microbes in China: 2009–2018. *Mar. Drugs* 17, 339. doi:10.3390/md17060339
- Terlouw, B. R., Blin, K., Navarro-Muñoz, J. C., Avalon, N. E., Chevrette, M. G., Egbert, S., et al. (2023). MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Res.* 51, D603–D610. doi:10.1093/nar/gkac1049
- Unno, R., Michishita, H., Inagaki, H., Suzuki, Y., Baba, Y., Jomori, T., et al. (1997). Synthesis and antitumor activity of water-soluble enediyne compounds related to dynemicin A. *Bioorg. Med. Chem.* 5, 987–999. doi:10.1016/s0968-0896(97)00037-0
- van Heel, A. J., de Jong, A., Song, C., Viel, J. H., Kok, J., and Kuipers, O. P. (2018). BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res.* 46, W278–W281–W281. doi:10.1093/nar/gky383
- Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Res.* 5, 737–746. doi:10.1101/gr.214270.116
- Wang, J.-T., Shi, T.-T., Ding, L., Xie, J., and Zhao, P.-J. (2023). Multifunctional enzymes in microbial secondary metabolic processes. *Catalysts* 13, 581. doi:10.3390/catal13030581
- Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., et al. (2016). Sharing and community curation of mass spectrometry data with global natural products social molecular networking. *Nat. Biotechnol.* 34 (8), 828–837. doi:10.1038/nbt.3597
- Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017a). Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb. Genom* 10, e000132. doi:10.1099/mgen.0.000132
- Wick, R. R., Judd, L. M., Gorrie, C. L., Holt, K. E., Unicycler, (2017b). Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* 13 (6), e1005595. doi:10.1371/journal.pcbi.1005595
- Wick, R. R., Judd, L. M., and Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* 1, 129. doi:10.1186/s13059-019-1727-y
- Wilke, D. V., Jimenez, P. C., Branco, P. C., Rezende-Teixeira, P., Trindade-Silva, A. E., Bauermeister, A., et al. (2021). Anticancer potential of compounds from the Brazilian blue amazon. *Planta Med.* 87, 49–70. doi:10.1055/a-1257-8402
- Xu, M., and Wright, G. D. (2019). Heterologous expression-facilitated natural products' discovery in actinomycetes. *J. Ind. Microbiol. Biotechnol.* 3–4, 415–431. doi:10.1007/s10295-018-2097-2
- Yadav, G., Gokhale, R. S., and Mohanty, D. (2009). Towards prediction of metabolic products of polyketide synthases: an *in silico* analysis. *PLOS Comput. Biol.* 5, e1000351. doi:10.1371/journal.pcbi.1000351
- Yamanaka, K., Reynolds, K. A., Kersten, R. D., Ryan, K. S., Gonzalez, D. J., Nizet, V., et al. (2014). Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci. U. S. A.* 5, 1957–1962. doi:10.1073/pnas.1319584111
- Yan, S., Zeng, M., Wang, H., and Zhang, H. (2022). *Micromonospora*: a prolific source of bioactive secondary metabolites with therapeutic potential. *J. Med. Chem.* 13, 8735–8771. doi:10.1021/acs.jmedchem.2c00626