



# The undiscovered natural product potential of *Actinomycetes*

Andrés M. Caraballo-Rodríguez<sup>1,2</sup> · Andrés Cumsille<sup>3,4</sup> · Sarolt Magyari<sup>5</sup> · Maria Taboada-Alquerque<sup>6</sup> · Bahar Behsaz<sup>7</sup> · Tiago F. Leão<sup>8</sup> · Kirk Broders<sup>9</sup> · Yasin El Abiead<sup>10</sup> · Jason A. Clement<sup>10</sup> · Vincent Charron-Lamoureux<sup>11</sup> · Simone Zuffa<sup>12</sup> · Louis-Félix Nothias<sup>13</sup> · Mengzhou Hu<sup>14</sup> · Christopher Leone<sup>1</sup> · Sarvar A. Kakhkhorov<sup>15</sup> · Beatriz Cámara<sup>3</sup> · Hosein Mohimani<sup>14,15</sup> · Pieter C. Dorrestein<sup>1,2,16,17</sup>

Received: 20 March 2025 / Revised: 2 October 2025 / Accepted: 15 October 2025 / Published online: 2 December 2025

© The Author(s) 2025. This article is published with open access

## Abstract

*Actinomycetes* have been a cornerstone species for the discovery of bioactive natural products with applications in pharmacotherapy and biotechnology. To expand the experimental evidence of their biosynthetic potential, we collected liquid-chromatography mass spectrometry untargeted metabolomics data on 948 microbial strains, mostly from *Actinomycetes*. This resulted in nearly two million MS/MS spectra, with an annotation rate of 13.3% corresponding to 2352 annotated molecules. Despite the efforts to link biosynthetic gene clusters to known molecules, most remain uncharacterized. This highlights the need for metabolomic data to bridge the gap between genomic potential and metabolite production. Although many unannotated spectra might correspond to different ion forms of the same molecule, the large amount of unknown molecules present in these datasets indicates that a significant number of natural products remain to be discovered, even within one of the most thoroughly studied sets of organisms. We provide a large metabolomics dataset as a public resource for data mining of microbial molecules and highlight its value by demonstrating the detection of edapochelins, recently discovered non-ribosomal peptides.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41429-025-00876-x>.

✉ Andrés M. Caraballo-Rodríguez  
acaraballorodriguez@health.ucsd.edu

<sup>1</sup> Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, San Diego, CA, USA

<sup>2</sup> Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, San Diego, CA, USA

<sup>3</sup> Departamento de Química y Centro de Biotecnología Daniel Alkalay Lowitt, Laboratorio de Microbiología Molecular y Biotecnología Ambiental, Universidad Técnica Federico Santa María, Valparaíso, Chile

<sup>4</sup> Department of Plant Pathology and Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, WI, USA

<sup>5</sup> Institute of Microbiology, Eidgenössische Technische Hochschule (ETH) Zürich, Zürich, Switzerland

<sup>6</sup> School of Pharmaceutical Sciences, University of Cartagena, Cartagena, Colombia

<sup>7</sup> Carnegie Mellon University & Chemia Biosciences Inc, Pittsburgh, PA, USA

<sup>8</sup> Núcleo de Bioensaios, Biossintese e Ecofisiologia de Produtos Naturais (NuBBE), Institute of Chemistry, São Paulo State University (UNESP), Araraquara, São Paulo, Brazil

<sup>9</sup> USDA, Agricultural Research Service, National Center for Agricultural Utilization Research, Mycotoxin Prevention and Applied Microbiology Research Unit, Peoria, IL, USA

<sup>10</sup> Baruch S. Blumberg Institute, Doylestown, PA, USA

<sup>11</sup> Institut de Chimie de Nice, Université Côte d'Azur CNRS, Nice, France

<sup>12</sup> Interdisciplinary Institute for Artificial Intelligence (3iA), Côte d'Azur, Sophia-Antipolis, France

<sup>13</sup> Laboratory of Physical and Chemical Methods of Research, Center for Advanced Technologies, Tashkent, Uzbekistan

<sup>14</sup> Carnegie Mellon University, Pittsburgh, PA, USA

<sup>15</sup> Department of Computational Medicine, University of California Los Angeles, Los Angeles, CA, USA

<sup>16</sup> Department of Pharmacology, University of California San Diego, San Diego, CA, USA

<sup>17</sup> Center for Microbiome Innovation, University of California San Diego, San Diego, CA, USA

## Introduction

*Actinomycetes*, in particular the *Streptomyces* genus, continuously proves to be prolific producers of unique bioactive molecules [1–3]. The biosynthetic potential of the *Streptomyces* genus has been demonstrated through genome mining approaches [4–6], revealing many biosynthetic gene clusters (BGCs) encoding the production of small molecules, also known as natural products. Despite being one of the most extensively studied genera from a natural products' standpoint, only an estimated 3% of their biosynthetic potential has been experimentally characterized [6]. Public deposition of sequenced genomes has accelerated the discovery of BGCs thanks to the development of genome mining tools such as antiSMASH [7], community efforts to annotate BGCs through initiatives such as MiBIG [8], and resources to link molecules to genes, such as the Paired Omics Data Platform [9]. However, public deposition of curated microbial metabolomes or metabolomics datasets, which could help to confirm the production of such molecules, are just starting [10, 11]. Although thousands of microbial molecules can be found in databases, such as NPAtlas [12] or Dictionary of Natural Products [13] only a few contain reference spectral data such as tandem mass spectrometry (MS/MS).

Several strategies are used to characterize the biosynthetic potential of microorganisms. On one hand, rapid advances in sequencing technologies coupled to genome mining tools have contributed to the understanding of the biosynthetic potential of microorganisms and even made possible analysis at the pan-genome scale [14]. However, limitations of these tools - due to inaccurate, incomplete or even lack of training data for developing new algorithms - can restrict the advance of knowledge. On the other hand, experimental data to validate the products of such BGCs is limited. Classical approaches involving culturing, extraction, isolation, purification and structural characterization are time consuming and expensive, however they have generated reliable data to confirm the expression of reported BGCs. These culture-based approaches involve testing different conditions on a single microbial strain [15], leading to discovering optimal factors that trigger the expression of silenced BGCs [16], further increasing interest in more in-depth biosynthetic studies.

Besides the advances in computational tools and resources available for metabolomics analysis, the lack of public metabolomics datasets remains a bottleneck to experimentally demonstrate the biosynthetic potential of microorganisms. Development of tools for predicting and comparing experimental data, as well as reference databases, has only been possible due to the availability of such public data. Additionally, well-characterized datasets are fundamental for further development in machine learning algorithms and benchmarking tools [17]. Such datasets will

enable the scientific community to further advance computational approaches, tools and resources while validating the biosynthetic potential of microorganisms.

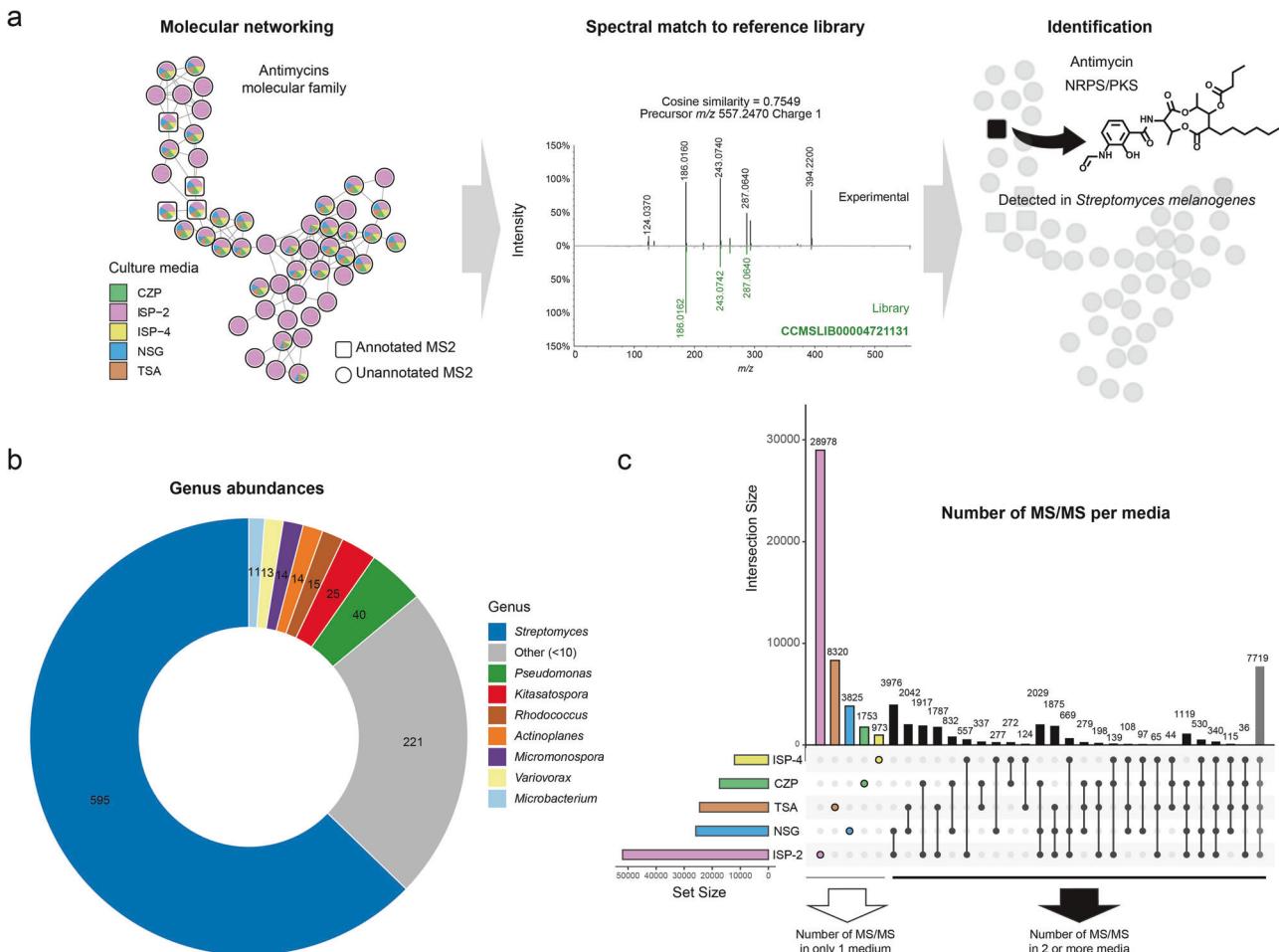
Due to the evident need for publicly available microbial metabolomes and as a resource to assess the discovery potential of small molecules, we provide metabolomics datasets for 948 microbial strains. We cultured 440 strains of *Actinomycetes* in five culture media to increase the success in production and detection of natural products. We expanded this analysis to include an additional 508 bacterial isolates cultured under a single condition (that also includes non-*Actinomycetes*, Supplementary Table S1), in order to increase the chemical diversity. Crude extracts were analyzed using untargeted liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). From these datasets, 260,488 out of the ~2 million detected MS/MS (~13.3% corresponding to 2352 molecules) were annotated using public spectral libraries. Therefore, this dataset serves as a valuable resource for microbial molecules data mining as demonstrated by detection of recently discovered non-ribosomal peptides and analogues.

## Results

### Detecting microbial molecules from 948 bacterial strains

Multiple factors influence the production of microbial small molecules. Strategies optimizing growth conditions and small molecules' production continue to evolve as our understanding of the environmental and chemical cues that influence microbial metabolism remains incomplete [18, 19]. Most strains in this study were sourced from a microbial collection with known optimal growth conditions. Therefore, we selected five complex media, including at least one optimal media for growth (see *Materials and Methods* section), to increase the success of detecting biosynthetic products [15].

Although there is a trade-off between microbial growth and detection of microbial molecules due to the interference of complex media and compatibility with mass spectrometry [18], the lack of available data and reference molecules to facilitate identification slows the discovery process. To expand the number of microbial molecules with available MS/MS data, we first studied the metabolome of 440 *Actinomycetes* using five culture media (Supplementary Fig. S1) and applied the molecular networking approach [20, 21]. Briefly, as shown in Fig. 1a, fragmentation spectra (MS/MS) are represented by nodes, connected based on spectral similarity and matched against spectral libraries for identification [21]. The number of MS/MS reflects the diversity of ion species, including adducts, in source fragments, multimeric species, rather than unique molecules



**Fig. 1** Production of small molecules from 948 bacterial strains. **a** Molecular networking approach provides a visualization of detected molecules (nodes represent detected molecules based on their fragmentation spectra MS/MS and connected based on spectral similarity). Color map indicates the culture media used in the study. Pie charts in the molecular family represent the proportion of fragmentation spectra MS/MS per media). Spectral match to reference library is a step performed after the molecular networking approach where each fragmentation spectra is compared against spectral libraries. Matches to reference spectra accelerate the identification of molecules. The mirror plot comparison of fragmentation spectra corresponds to antimycin, a depsipeptide assembled from a hybrid non-ribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) biosynthetic pathway. Experimental MS/MS (top, black) and reference spectra from GNPS spectral libraries (bottom, green). This identification allows to annotate the molecular family as antimycins as connected nodes share chemical relationships based on spectral similarity; **b** Diversity of the bacterial

[22–35]. However, a higher number of observed MS/MS spectra under a given culturing condition increases the likelihood of capturing a greater diversity of unique molecules. Therefore we use the number of MS/MS spectra as a proxy to assess culturing conditions that maximize the number of observed molecules.

Among the five media used, ISP-2 (International *Streptomyces* Project medium 2) [36] has been commonly used for the growth, production and detection of small molecules

strains based on genus. *Streptomyces* genus forms the highest proportion of the selected strains for this study. The “Other (<10)” category refers to genera with fewer than 10 strains. This category comprises 221 strains, belonging to 102 genera; c Upset plot showing the number of fragmentation spectra (MS/MS) per culture condition (culture media). In this visualization, the Intersection Size indicates the number of MS/MS that are unique (colored according to each culture condition), shared between two or more culture conditions (black) and shared among all culture media (grey). These are shown in the matrix and in the top bar-plot. The Set Size visualized as the left bar-plot corresponds to the total number of fragmentation spectra (MS/MS) per individual culture condition, regardless of their overlap with other culture conditions. These are shown in the matrix and in the left bar-plot. In short, the Set Size bars indicate how many MS/MS are per culture condition individually, while the intersection bars indicate how many MS/MS are in overlap with other culture media

[18, 36]. Additionally, ISP-2 medium has shown compatibility with mass spectrometry approaches without producing high background or interference [37]. We observed that for some molecular families, such as the depsipeptides antimycins [38], most of the MS/MS were present in this media (Fig. 1a), indicating that ISP-2 still supported the detection of a considerable number of molecules (Supplementary Fig. S1). Therefore, due to ISP-2 being commonly and successfully used for both growing and production of

microbial molecules, we included in the analysis additional isolates mostly corresponding to the genus *Streptomyces* (Fig. 1b) cultured using this medium. It is important to highlight that a high number of unique MS/MS were detected from each medium, including ISP-4, which was the medium with the lowest number of unique MS/MS (973 spectra) (Fig. 1c). This is consistent with the OSMAC strategy [15], in which variation of growing conditions, such as culture media, has a direct effect on the levels of production and diversity of small molecules. Additionally, a substantial number of detected MS/MS were shared across all five media (7719 detected MS/MS, Fig. 1c), highlighting their suitability for microbial molecule production. The annotation rate (~13.3%) of the data generated through LC-MS/MS suggests significant potential for discovery of microbial natural products. We employed untargeted LC-MS/MS data to overcome the potential interference of ion suppression observed using direct infusion methods without chromatography to reduce chemical complexity during detection by MS. The metabolomics data is publicly available to facilitate future discovery of microbial molecules using spectral searches within the GNPS environment or using other tools including MZmine and MS-DIAL [39, 40].

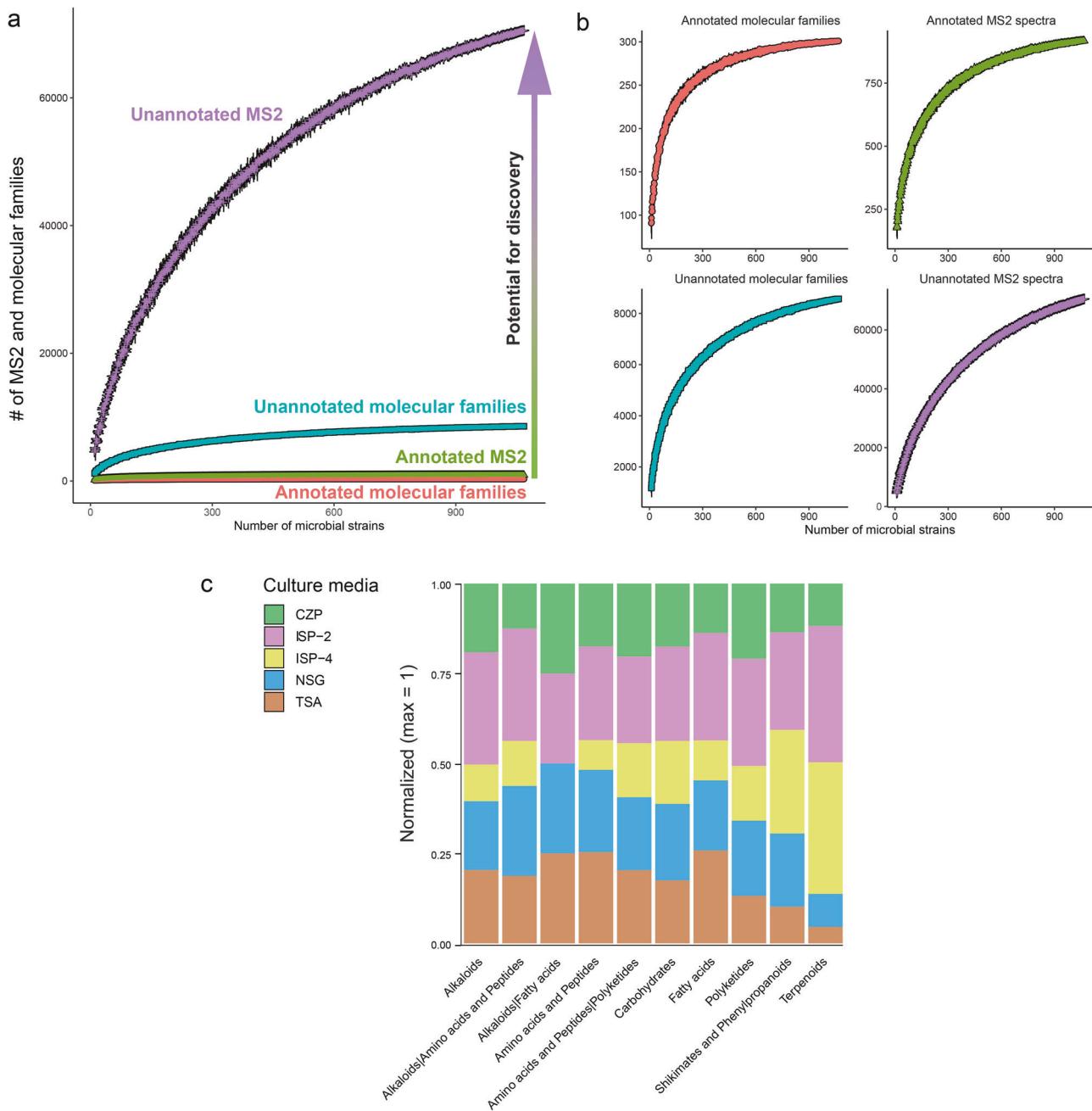
## Annotating the data

The entire dataset of 948 strains contains a total of 1,947,363 fragmentation spectra that represent 105,499 unique MS/MS spectra after MS-Cluster merging. Molecules from the culture media were removed by using exclusion lists during data acquisition (see Materials and Methods section), while signals from noise were filtered out during molecular networking using blank samples (see Materials and Methods section, Molecular Networking settings). Although the number of natural products will be far lower since one molecule might be found in many different ion forms due to in-source fragmentation, different adducts or multimeric species, this represents the upper bound of the discovery potential of candidate molecules that is present in this data set. This is important, as different researchers might be able to access different instruments and methods for LC-MS/MS, and detection of molecules as different species might facilitate the chances of discovery of molecules within these datasets. We used existing public MS/MS libraries in the Global Natural Products Social molecular networking ecosystem (GNPS) to provide annotations [20, 21]. This resource includes more than 30 different public reference libraries (listed in the Materials and Methods section).

GNPS also has propagated libraries [41]. These are libraries where the MS/MS is annotated based on similarity to another annotated MS/MS spectrum. This approach

enables the discovery of analogs and metabolized versions of related molecules. When combining experimental and propagated libraries [41], we were able to double the number of annotated spectra from 1235 (130,317 MS/MS spectra) to 2352 (260,488 MS/MS spectra, see networking links provided in the Materials and Methods section). It should be noted that not every annotation will be correct. The annotation settings used give rise to ~1% false discovery rates [42].

Examples of annotated MS/MS spectra (Fig. 1) include the antimycin antibiotics. These molecules are depsipeptides assembled from a hybrid non-ribosomal peptide synthetase (NRPS)/polyketide synthase (PKS) biosynthetic pathway encoded by *Streptomyces* genus [43, 44]. Information regarding geographical distribution of the *Streptomyces* isolates used in this study enabled to confirm previous findings indicating a worldwide distribution of antimycin BGCs [44]. Other examples of bioactive detected molecules include actinomycins, antibiotic and anticancer compounds [45–48]; pteridic acid, a polyketide produced by *Streptomyces* and responsible for plant resistance to abiotic stress [49]; sevadincin, a NRP previously reported from *Paenibacillus larvae* [50]; staurosporine, an apoptosis inducer alkaloid originally isolated from *Streptomyces* [51, 52]; trienomycin, antibiotic with anticancer potential [53, 54]; spiramycin, a long time used macrolide antibiotic in humans originally isolated from soil *Streptomyces* [55]; josamycin (also known as leucomycin A3, kitasamycin A3, turimycin A5) a macrolide antibiotic with antiviral properties against influenza [56]; the anticancer agent daunorubicin, a member of the type II polyketide family [57]; the immunosuppressant agent tacrolimus (also known as FK-506), a polyketide macrolide more potent than cyclosporin [58, 59]; and, siderophores, such as desferrioxamine [60], a chelating agent used to remove excess metals with therapeutic applications for humans (e.g., desferrioxamine B, commercially known as Desferal) [61]. After manual curation of the experimental data, we provided 1215 annotated spectra corresponding to 698 non-redundant microbial molecules (Supplementary Table S2) created from classical molecular networking and library annotations in GNPS2 (link provided in Materials and Methods section). We expanded the annotations to 2352 by using propagation libraries [41]. Additionally, we created a database of 515 known molecules from *Streptomyces* (Supplementary Table S3), their fragmentation spectra were predicted and compared against the experimental data acquired for this study. Then, manual verification enabled us to confirm 116 molecules (Supplementary Table S4), providing additional annotations to the ones obtained through the GNPS public libraries. The database corresponding to the 515 known molecules from *Streptomyces* is available as Supplementary Table S3, which can be used to propagate annotations

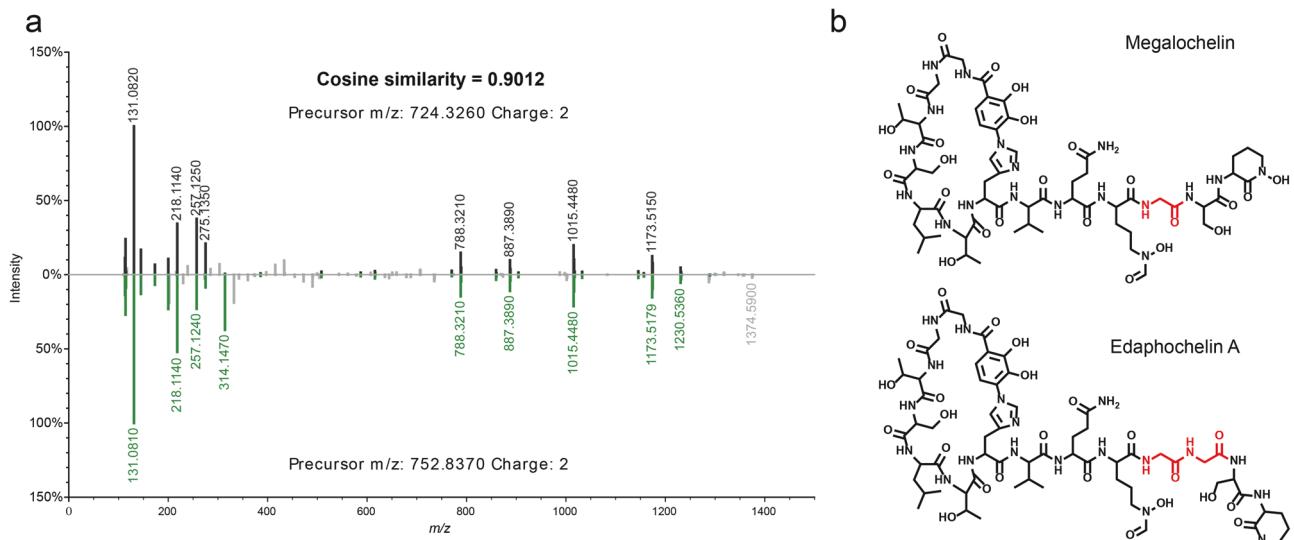


**Fig. 2** Chemical space detected from microbial strains in this study. **a** Rarefaction curve displaying the number of strains (x-axis) against the number of MS/MS spectra and molecular families (y-axis); **b** Zoom in of rarefaction curves by annotated molecular families, annotated MS/MS spectra, unannotated molecular families and unannotated MS/MS spectra. Color code as shown in the figures; **c** Number of annotated spectra using GNPS spectral libraries and classified by

biosynthetic pathways (NPclassifier) [65]. Bar plots showing the variation of the annotated molecules by five of the culture media. The number of annotations as shown in Fig. 1 (main text) was normalized to 1 as maximum by biosynthetic pathway. Briefly, the number of annotated MS/MS was calculated per culture condition and per biosynthetic pathway and normalized to the maximum within each pathway. Color code as shown in the figure per culture condition

[62, 63], to match experimental data of molecules that have not been deposited in public libraries of experimental data until now. The curated annotations were uploaded to the Collaborative Microbial Metabolite Center CMMC (<https://cmmc.gnps2.org/>) and are publicly available.

This resource adds reference spectra for more than a thousand known molecules, including membrane lipids and other common natural product metabolites, such as those nutrients (e.g., amino acids) present in culture media. Then, even if no annotation is provided, its connection to other



**Fig. 3** Recently discovered siderophores from *Streptomyces* sp. **a** Mirror comparison of MS/MS spectra of megalochelin [68] (top) and edaphochelin A (Behsaz et al. submitted) (bottom); **b** Chemical structures of megalochelin (top) and edaphochelin A (bottom). The

location of the additional glycine residue predicted by ModiFinder [69] (Supplementary Fig. S3) is consistent with the chemical structure of edaphochelin A

similar spectra within a molecular family increases the probability of annotation at the molecular family (e.g., propagated annotation) as shown for antimycins in Fig. 1a. Our work increases the experimental evidence of the Actinomycetota metabolome, mostly from *Streptomyces* species. Therefore, we provided proof of detection and curated annotations for more than thousand microbial molecules produced by the selected bacteria strains included in this study.

### Estimating the discovery potential

To estimate the discovery potential, we used a rarefaction approach (Fig. 2a, b, see Methods for details). Using rarefaction, each MS/MS spectrum in the data of a sample is counted, then for the next sample each additional MS/MS that is not found in the data from the first sample is counted, and then for the third sample, data will be added but only if the MS/MS is not found in the first two samples. Rarefaction is continued in this fashion until all the MS/MS are found in the data from the 948 bacterial samples (3439 files). This can be performed after merging all identical spectra (consensus spectra) and also at the molecular family level (connected nodes). This analysis demonstrates that most MS/MS spectra remain unannotated. However, by grouping these spectra into molecular families, the annotation gap is reduced to provide a higher level of the detected chemical diversity and potential of discovery. The output of GNPS based molecular networking, which includes a clustering step performed by MS-Cluster [21, 64], was used to perform the rarefaction analysis. The rarefaction analysis

reveals that the number of annotated molecular families is 328 while the unannotated families are reaching above 8000, highlighting that there is a vast number of molecular families that remain unannotated. This includes unannotated families of different ion forms of the annotated molecular families where the MS/MS does not share sufficient similarity to known ions. As molecular networking is complemented with search against spectral libraries and further chemical classification is possible, additional evidence regarding a chemical diversity approach was provided (Fig. 2c, Supplementary Fig. S2) based on chemical classification using NPClassifier [65]. To facilitate the comparison across different culture conditions, the number of annotated spectra was normalized per culture condition and biosynthetic pathway (Fig. 2c). It was observed that ISP-2 provided a more consistent detection of molecules across the biosynthetic pathways and in some cases, relatively higher. This was evident when comparing ISP-2 and TSA regarding the biosynthetic pathway of terpenoids. In other cases, for instance when focusing on ISP-4, low proportion of annotations belonging to Alkaloids and Alkaloids/Amino acids and Peptides were observed, while no detection of molecules belonging to Alkaloids/Fatty acids pathway was observed. Altogether, these observations further supported the inclusion of additional datasets corresponding to microbial extracts obtained in ISP-2 to increase the coverage of detected molecules from this study. This is consistent with other reports indicating high antimicrobial activity of *Streptomyces* strains when cultured on ISP-2 media [66, 67].

As a demonstration of the potential of discovery that the datasets provided in this work contain, a recently discovered

NRP siderophore related to megalochelin [68] was detected. This analogue of megalochelin was also detected from a *Streptomyces* strain and identified as edaphochelin A (Behsaz et al., submitted). Since both molecules shared a chemical similarity based on their fragmentation pattern (Fig. 3), the chemical structure of megalochelin was used to confirm such similarity and more importantly, the location of the chemical modification observed in edaphochelin A. By using ModiFinder [69], a recently developed tool able to predict the location of a chemical modification of an unknown molecule based on its spectral similarity when compared against a known molecule, an additional glycine residue was predicted to be present in edaphochelin A (Supplementary Fig. S3). The prediction provided by ModiFinder is consistent with the chemical structure proposed for edaphochelin A and a remarkable example that showcases the use of metabolomics datasets of this kind and application of computational developments to further expand the discovery of microbial molecules present in public repositories.

## Discussion

Several gaps still exist in the search for new microbial small molecules. While tens of thousands of microbial molecules have been discovered to date [12], the genetic information, in the form of gene clusters, encoded in microbial genomes suggests a potential exceeding millions [6]. However, genetic potential does not guarantee production under laboratory conditions. Therefore, detecting the products of such genomic potentials opens several avenues for investigating the roles of such molecules mediating interkingdom relationships, such as plant-microbe interactions (e.g., pteridic acids reverse drought and salinity stress in plants) [49]. In addition, further biosynthetic studies are facilitated once a gene product has been detected to demonstrate the geographical distribution and evolution of BGCs (e.g., pteridic acids and antimycins BGCs distribution) [44, 49]. The discovery of BGCs is expected, and we envision many of the detected molecules in our datasets will be linked to such microbial genes. A recently discovered group of NRP siderophores, megalochelins [68], that show a characteristic fragmentation pattern also shared by edaphochelins (Behsaz et al., submitted) (Fig. 3) highlight the value of publicly available experimental microbial metabolomics datasets such as the one provided in this study to accelerate the discovery and link to BGCs of microbial producers. By considering estimates, where it was suggested that one *Streptomyces* genome may contain between 8–83 BGCs [70], and could produce 100,000 molecules [71, 72], access to public datasets for mining microbial molecules is necessary. Since a single BGC can be involved in the

biosynthesis of multiple compounds [73], the number of microbial molecules is expected to far exceed the number of BGCs present in a genome.

Current strategies for identifying microbial molecules rely on publicly available chemical information in spectral libraries and databases. By using tandem mass spectrometry, the need for reference spectra to assist the dereplication steps is evident. However, the lack of such references causes delays in the discovery of new molecules, a key point in the search for new antibiotics and other bioactive molecules. This is one of the main reasons for parallel strategies by propagating annotation of tandem mass spectrometry data [62, 74, 75]. For instance, the recent discovery of angucycline analogues [76], one of the largest families of polyketides produced by *Streptomyces*, was possible due to the annotation of detected spectra from previous work [77], and their availability in public libraries, such as GNPS spectral libraries. To accelerate the discovery of new microbial molecules, efforts to increase the amount of reference spectra must be made.

In addition to providing a large and curated microbial reference dataset, we have made these datasets public and accessible as part of community resources via the GNPS platform and the Mass Spectrometry Interactive Virtual Environment (MassIVE) repository (See Data availability section). These datasets have also been integrated in microbeMASST [78], a novel MS/MS search tool, representing an evolution of the original MASST [79], that allows searching for single MS/MS spectra against a reference microbial metabolomics database. With this tool, researchers can search unknown fragmentation spectra and gain information about microbial producers, confirming or generating biosynthetic hypotheses. Furthermore, we have provided curated annotations as part of the Collaborative Microbial Metabolite Center CMMC (<https://cmmc.gnps2.org/>). This large dataset is also part of the community effort to enable metabolomic analysis at the repository scale via ReDU (<https://redu.gnps2.org/>) [80, 81]. In line with the need for data standardization, which includes growth media and culture conditions [82], this resource supports a wide range of applications, from identifying small molecule producers to benchmarking annotation tools and improving predictive algorithms for linking molecules to genes.

## Conclusions

We provided tandem mass spectrometry data from hundreds of bacterial strains, primarily from *Streptomyces* genus. We provided 1215 annotated spectra corresponding to 698 non-redundant microbial molecules in addition to 116 annotations based on propagated annotations that had not been annotated yet in public mass spectrometry libraries. Our results will

support future research focused on investigating the ecological role of thousands of molecules detected in our datasets, even though it is well known that many BGCs are silent or cryptic under laboratory conditions. Additionally, many more predicted BGCs within other uncovered microbial groups are likely to encode the reported produced compounds with relevant ecological and pharmacological activities. Finally, we envision this large dataset of microbial molecules and associated information as a valuable resource, not just for dereplication purposes but also, to identify unknown producers of known bioactive molecules ultimately facilitating the linkage of molecules to their biosynthetic genes.

### Limitations/scope

The detected molecules are limited to the culture conditions, extractions, liquid chromatography and mass spectrometry settings used during the data acquisition. Several fragmentation spectra might correspond to adducts, in source fragments and multimers belonging to the same molecule, so redundancy is expected in the data set. Yet, despite those caveats, it shows there is a lot of discovery potential. Annotation of fragmentation spectra enables us to begin to tackle this redundancy, but annotation rates are still limited to available spectral libraries as stated in this study.

## Materials and methods

### Materials

*Actinomycetes* strains were obtained from the Agricultural Research Service Culture Collection (NRRL) <https://nrrl.ncaur.usda.gov/>.

Nunc™ 96-Well Polypropylene DeepWell™ storage plates, catalog number: 278743; acetonitrile with formic acid 0.1% (v/v) LC-MS grade (Optima, Fisher Chemical) catalog no. LS120-4; water with formic acid 0.1% (v/v) LC-MS grade (Optima, Fisher Chemical) catalog no. LS118-4.

### Actinomycetes strains

Microorganisms received from the provider as freeze dried pellets were resuspended in 50% glycerol and transfer to 96 deep well plates for inoculation of media in 96-DeepWell plates and stored under -80 °C until microbial cultures were initiated.

### Media

All media were prepared according to the provider recommendations and recipes available through the NRRL (<https://nrrl.ncaur.usda.gov/>) for each *Actinomycetes* strain:

International *Streptomyces* Project Yeast Extract-Malt Extract agar (ISP-2 agar) (Difco No. 277010); International *Streptomyces* Project Synthetic Salts-Starch Medium (ISP-4 agar) (Difco No. 277210); Trypticase Soy Agar (TSA) (Difco No. 211768); N-Z Amine with Soluble Starch and Glucose agar (NSG agar) (Sigma Chemical Company Catalog No. C 0626); Czapek's Solution and agar (CZP agar) (Sigma Aldrich Catalog No. C 6095).

### Culture and incubation

Cultures of microbial strains were performed as follows: from glycerol stocks, an inoculum of 20  $\mu$ L of glycerol stocks were transferred into 96-deep well plates containing 1 mL of solid media and incubated at 28 °C during 14 days.

### Extraction

After a 14-day incubation period, plates were placed under -80 °C until extraction procedure. Three freeze-thaw cycles of 10 min each were performed before extraction steps. For agar-based media cultures, 1 mL of methanol was added to each well, and plates were submitted to sonication for 15 min (Branson 5510, Marshall Scientific, Hampton, NH, USA), centrifugation for 15 min at 2000 RPM (865  $\times$  g) (Sorvall Legend RT, Marshall Scientific, Hampton, NH, USA), transferring of supernatant to a clean 96-well plate and dried out in a centrifugal vacuum concentrator, CentriVap (Labconco, Kansas City, MO, USA). Samples were dissolved in 500  $\mu$ L 50% methanol:water LC-MS grade, and 100  $\mu$ L were transferred to a clean 96-well plate and dried out in the centrifugal vacuum concentrator.

### LC-MS/MS acquisition

Obtained extracts were dissolved in 200  $\mu$ L 50% methanol<sub>(aq)</sub> containing 1  $\mu$ M sulfadimethoxine as internal standard for LC-MS monitoring. Untargeted LC-MS/MS acquisition was performed on a Vanquish Ultrahigh Performance Liquid Chromatography (UPLC) system coupled to a Q-Exactive™ Hybrid Quadrupole-Orbitrap™ (Thermo Fisher Scientific, Bremen, Germany). Chromatographic separation was performed on a Kinetex 1.7  $\mu$ m 100 Å pore size C18 reversed phase UHPLC column 50  $\times$  2.1 mm (Phenomenex, Torrance, CA) with a constant flow rate of 0.5 mL/min. The following solvents were used during the LC-MS/MS acquisition: water with 0.1% formic acid (v/v), Optima™ LC/MS Grade, Thermo Scientific™ (solvent A) and acetonitrile with 0.1% formic Acid (v/v), Optima™ LC/MS Grade, Thermo Scientific™ (solvent B). After injection of 2  $\mu$ L of sample into the LC system, the elution was performed isocratically with 5% B from 0 to 0.5 min, then with a multistep linear gradient from 5 to 50% B (0.5–6 min), 50 to 99% B (6–8 min), 99% B

(8–10 min), 99 to 5% B (10–10.5 min), and 5% B (10.5–12 min). Data dependent acquisition (DDA) mode was used for acquisition of tandem MS (MS/MS) data with a default charge state of 1. An exclusion list was generated for each culture media and used as input for untargeted LC-MS/MS in data dependent acquisition mode. The exclusion list was created using a Python script available in: [https://github.com/lfnothias/IODA\\_MassSpec](https://github.com/lfnothias/IODA_MassSpec) [83]. Full MS was acquired using 1 microscan at a resolution (R) of 35,000 at  $m/z$  200, automatic gain control (ACG) target 5e5, maximum injection time (IT) of 100 ms, scan range  $m/z$  250–3750 and data acquired in profile mode. DDA of MS/MS was acquired using 1 microscan at a resolution (R) of 35,000 at  $m/z$  200, automatic gain control (ACG) target 5e5, top 5 ions selected for MS/MS with isolation window of  $m/z$  2.0 with scan range  $m/z$  200–2000, fixed first mass of  $m/z$  100 and stepped normalized collision energy (NCE) of 20, 25 and 30 eV, minimum ACG target 2.50e4, intensity threshold 2.5e5, apex trigger 2–15 s, all multiple charges included, isotopes were excluded and a dynamic exclusion window of 5 s. Analytical blanks and mixture of sulfamethazine, sulfamethizole, sulfachloropyridazine, sulfadimethoxine, amitriptyline, and coumarin-314 (10  $\mu$ M) were injected after every 96 samples as quality control for monitoring instrument (LC-MS) performance.

## Molecular networking

A molecular network was created using the online workflow (<https://ccms-ucsd.github.io/GNPSDocumentation/>) on the GNPS website (<http://gnps.ucsd.edu>) [20]. The data was filtered by removing all MS/MS fragment ions within  $\pm 17$  Da of the precursor  $m/z$ . MS/MS spectra were window filtered by choosing only the top 6 fragment ions in the  $\pm 50$  Da window throughout the spectrum. The precursor ion mass tolerance was set to 0.02 Da and a MS/MS fragment ion tolerance of 0.02 Da. A network was then created where edges were filtered to have a cosine score above 0.7 and more than 6 matched peaks. Further, edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. Finally, the maximum size of a molecular family was set to 100, and the lowest scoring edges were removed from molecular families until the molecular family size was below this threshold. The spectra in the network were then searched against GNPS' spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least 6 matched peaks.

The molecular networking analysis with metadata including culture media can be accessed through the link: <https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=aa06c41deb25441b818f5e716dd1d095>.

The molecular networking analysis with additional metadata including culture media without removing MS/MS from blanks can be accessed through the link: <https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=f69176f25e3d4aed93f335b05b8a5fba>.

The molecular networking analysis for 440 *Actinomycetes* with additional metadata including culture media without removing MS/MS from blanks (blanks already removed using exclusion list during LC-MS/MS acquisition) can be accessed through the link: <https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=cb82ed06161046f5b9cd9f8638374eea>.

Molecular networking was also created using the GNPS2 platform and maintaining the same settings described above. This facilitated the visualization of the large dataset composed of 4020 files, 593 corresponding to blanks and quality controls used to monitor instrument performance during LC-MS/MS acquisition. The molecular networking analysis can be accessed through the following link (this job is used for Supplementary Table S2): <https://gnps2.org/status?task=e4a5b96dddf4ccb8f7ab54095684e10>.

## In silico annotation via network annotation propagation

Network annotation propagation (NAP) [62] was performed on a subset of the dataset (Classical molecular networking job link: <https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=0ea632474af24eb6a84e576d0aa585ca>) due to memory limitations (running NAP on entire dataset is not currently possible, but improvement efforts are ongoing via Chem-Walker approach [63]), via the GNPS platform using a customized database containing 515 reported products of *Streptomyces* not yet available through GNPS libraries (Supplementary Table S3). The parameters used the 10 first candidates for consensus score, 15 ppm accuracy for exact mass candidate search, positive acquisition mode, and 0.5 cosine value to subselect inside a cluster. Fusion results were used to determine a consensus, searching only for  $[M + H]^+$  adduct type. A maximum of 10 candidate structures were used in the graph. The following databases were searched: Dictionary of Natural Products (DNP) [13], Super Natural II [84], GNPS [20], and Chemical Entities of Biological Interest (ChEBI) [85]. GNPS NAP job link: <https://proteomics2.ucsd.edu/ProteoSAFe/status.jsp?task=b89434115a5248d08b92ecb6d6a719d1>.

## Propagated annotation via suspects library

In addition to the spectral libraries included in the molecular networking analysis, propagated annotations were obtained on the same datasets using the suspects libraries [41]. The molecular networking analysis including the suspects library can be accessed through the link: <https://gnps.ucsd.edu>.

[edu/ProteoSAFe/status.jsp?task=22521ea20bf5412a819124fac1db0efb](https://edu/ProteoSAFe/status.jsp?task=22521ea20bf5412a819124fac1db0efb).

The following link provides access to the molecular networking analysis and library search with all libraries including propagated (e.g., suspects library) in the GNPS2 platform: <https://gnps2.org/status?task=d6a0bab81cd844bb8cddcef17764a32>.

## Data availability

All the datasets used and generated for this study have been deposited online in the public repository Mass Spectrometry Interactive Virtual Environment (MassIVE). Bacterial strains were provided by the NRRL Agricultural Research Service Culture Collection from the U. S. Department of Agriculture (USDA) and RIKEN BioResource Research Center. Public genome data from all bacterial strains used in this study are available in the NCBI genome database. The following mass spectrometry datasets can be accessed through the MassIVE repository at <https://massive.ucsd.edu/>: MSV000088235 GNPS - 100 *Actinobacteria* strains in ISP-2 media; MSV000088742 GNPS - 95 *Actinobacteria* strains in ISP-2 liquid media; MSV000088196 GNPS - *Populus* trees 153 microbial isolates; MSV000088763 GNPS - *Actinobacteria* cultured in ISP-2 agar media; MSV000088764 GNPS - *Actinobacteria* cultured in NSG agar media; MSV000088800 GNPS - *Actinobacteria* cultured in Czapek agar media; MSV000088801 GNPS - *Actinobacteria* cultured in TSA media; MSV000088816 GNPS - *Actinobacteria* cultured in ISP-4 media; MSV000089090 GNPS - 95 *Actinobacteria* strains in ISP-2 liquid media; MSV000089742 GNPS - 95 *Actinobacteria* strains in ISP-2 liquid media; MSV000089813 GNPS - *Actinobacteria* cultured in ISP-2 agar media; MSV000089815 GNPS - *Actinobacteria* cultured in NSG agar media; MSV000089816 GNPS - *Actinobacteria* cultured in Czapek agar media; MSV000089817 GNPS - *Actinobacteria* cultured in TSA media; MSV000089818 GNPS - *Actinobacteria* cultured in ISP-4 media.

**Acknowledgements** We thank Travis Adkins and Lisa McCormick from the USDA ARS Culture Collection for their assistance selecting and providing microbial strains and Dale Pelletier from Oak Ridge National Laboratory for providing microbial extracts used in this research. This project was supported in part by the U.S. Department of Agriculture, Agricultural Research Service. AMC-R and PCD were supported by the National Institutes of Health (NIH) grants 1DP2GM137413-01, 5R01GM107550-10, 1U24DK133658-01 and by the Gordon and Betty Moore Foundation, GBMF12120 and <https://doi.org/10.37807/GBMF12120>. AC acknowledges ANID 21191625 PhD fellowship and Programa de Incentivos a la Iniciación Científica, UTFSM, BC acknowledges Fondecyt regular project N° 1221264 and ANID-Milenio-NCN2023\_054. S.M. the ETH Zurich Doc. Mobility Fellowship and M.T. the BPIN 2020000100364 Minciencias PhD fellowship. SAK was supported by the Fund for Financing Science and Supporting Innovation under the Agency of Innovative Development

of the Republic of Uzbekistan. Support for title page creation and format was provided by AuthorArranger, a tool developed at the National Cancer Institute.

## Compliance with ethical standards

**Conflict of interest** There are no known conflicts of interest in this work by the USDA, Agricultural Research Service, National Center for Agricultural Utilization Research, Mycotoxin Prevention and Applied Microbiology Research Unit. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. PCD is an advisor and holds equity in Cybele, BileOmix and Sirenas and a Scientific co-founder, advisor, holds equity and/or received income to Ometa, Enveda, and Arome with prior approval by UC-San Diego. PCD also consulted for DSM animal health in 2023. HM and BB are co-founders and have equity interests from Chemia Biosciences Inc.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Kong D, Wang X, Nie J, Niu G. Regulation of Antibiotic Production by Signaling Molecules in *Streptomyces*. *Front Microbiol*. 2019;10. <https://www.frontiersin.org/articles/10.3389/fmicb.2019.02927>
2. Zhang Z, Du C, de Bary F, Liem M, Liakopoulos A, van Wezel GP, et al. Antibiotic production in *Streptomyces* is organized by a division of labor through terminal genomic differentiation. *Sci Adv*. 2020;6:eaay5781.
3. Donald L, Pipite A, Subramani R, Owen J, Keyzers RA, Taufa T. *Streptomyces*: Still the Biggest Producer of New Natural Secondary Metabolites, a Current Perspective. *Microbiol Res*. 2022;13:418–65.
4. Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, et al. Minimum Information about a Biosynthetic Gene cluster. *Nat Chem Biol*. 2015;11:625–31.
5. Leão TF, Wang M, da Silva R, Gurevich A, Bauermeister A, Gomes PWP, et al. NPOmix: A machine learning classifier to connect mass spectrometry fragmentation data to biosynthetic gene clusters. *PNAS Nexus*. 2022;1:pgac257.
6. Gavrililidou A, Kautsar SA, Zaburannyi N, Krug D, Müller R, Medema MH, et al. Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. *Nat Microbiol*. 2022;7:726–35.
7. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, et al. antiSMASH 6.0: improving cluster

detection and comparison capabilities. *Nucleic Acids Res.* 2021;49:W29–35.

- Terlouw BR, Blin K, Navarro-Muñoz JC, Avalon NE, Chevrette MG, Egbert S, et al. MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Res.* 2023;51:D603–10.
- Schorn MA, Verhoeven S, Ridder L, Huber F, Acharya DD, Aksенов AA, et al. A community resource for paired genomic and metabolomic data mining. *Nat Chem Biol.* 2021;17:363–8.
- Tay DWP, Tan LL, Heng E, Zulkarnain N, Chin EJ, Tan ZYQ, et al. Tandem mass spectral metabolic profiling of 54 actinobacterial strains and their 459 mutants. *Sci Data.* 2024;11:977.
- Simone M, Iorio M, Monciardini P, Santini M, Cantù N, Tocchetti A, et al. The Molecules Gateway: A Homogeneous, Searchable Database of 150k Annotated Molecules from Actinomycetes. *J Nat Prod.* 2024;87:2615–28.
- van Santen JA, Poynton EF, Iskakova D, McMann E, Alsup TA, Clark TN, et al. The Natural Products Atlas 2.0: a database of microbially-derived natural products. *Nucleic Acids Res.* 2022;50:D1317–23.
- Dictionary of Natural Products 31.1 Chemical Search. 2022. Available from: <https://dnp.chemnetbase.com/faces/chemical/ChemicalSearch.xhtml>
- Mohite OS, Jørgensen TS, Booth T, Charusanti P, Phaneuf PV, Weber T, et al. Pangenome mining of the *Streptomyces* genus redefines their biosynthetic potential. *bioRxiv.* 2024. <https://www.biorxiv.org/content/10.1101/2024.02.20.581055v1>
- Bode HB, Bethe B, Höfs R, Zeeck A. Big Effects from Small Changes: Possible Ways to Explore Nature's Chemical Diversity. *ChemBioChem.* 2002;3:619–27.
- Covington BC, Xu F, Seyedsayamdst MR. A Natural Product Chemist's Guide to Unlocking Silent Biosynthetic Gene Clusters. *Annu Rev Biochem.* 2021;90:763–88.
- Ebbels TMD, van der Hooft JJ, Chatelaine H, Broeckling C, Zamboni N, Hassoun S, et al. Recent advances in mass spectrometry-based computational metabolomics. *Curr Opin Chem Biol.* 2023;74:10228.
- Donnelly AE, Narayanan N, Biner-Williams CMC, DeWolfe TJ, Chu RK, Anderton CR, et al. Balancing Trade-Offs Imposed by Growth Media and Mass Spectrometry for Bacterial Exometabolomics. *Appl Environ Microbiol.* 2022;88:e00922-22.
- Behie SW, Bonet B, Zacharia VM, McClung DJ, Traxler MF. Molecules to Ecosystems: Actinomycete Natural Products In situ. *Front Microbiol.* 2017;7. <https://www.frontiersin.org/articles/10.3389/fmicb.2016.02149>
- Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol.* 2016;34:828–37.
- Aron AT, Gentry EC, McPhail KL, Nothias LF, Nothias-Esposito M, Bouslimani A, et al. Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nat Protoc.* 2020;15:1954–91.
- Bernardo-Bermejo S, Xue J, Hoang L, Billings E, Webb B, Honders MW, et al. Quantitative multiple fragment monitoring with enhanced in-source fragmentation/annotation mass spectrometry. *Nat Protoc.* 2023;18:1296–315.
- Brown M, Wedge DC, Goodacre R, Kell DB, Baker PN, Kenny LC, et al. Automated workflows for accurate mass-based putative metabolite identification in LC/MS-derived metabolomic datasets. *Bioinformatics.* 2011;27:1108–12.
- Chen L, Pan H, Zhai G, Luo Q, Li Y, Fang C, et al. Widespread occurrence of in-source fragmentation in the analysis of natural compounds by liquid chromatography–electrospray ionization mass spectrometry. *Rapid Commun Mass Spectrom.* 2023;37:e9519.
- El Abied Y, Bueschl C, Panzenboeck L, Wang M, Doppler M, Seidl B, et al. Heterogeneous multimeric metabolite ion species observed in LC-MS based metabolomics data sets. *Anal Chim Acta.* 2022;1229:340352.
- Giera M, Aisporna A, Uritboonthai W, Siuzdak G. The hidden impact of in-source fragmentation in metabolic and chemical mass spectrometry data interpretation. *Nat Metab.* 2024;6:1647–8.
- Kuhl C, Tautenhahn R, Böttcher C, Larson TR, Neumann S. CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets. *Anal Chem.* 2012;84:283–9.
- Mahieu NG, Spalding JL, Gelman SJ, Patti GJ. Defining and Detecting Complex Peak Relationships in Mass Spectral Data: The Mz.unity Algorithm. *Anal Chem.* 2016;88:9037–46.
- Mahieu NG, Patti GJ. Systems-Level Annotation of a Metabolomics Data Set Reduces 25 000 Features to Fewer than 1000 Unique Metabolites. *Anal Chem.* 2017;89:10397–406.
- Nash WJ, Ngere JB, Najdekr L, Dunn WB. Characterization of Electrospray Ionization Complexity in Untargeted Metabolomic Studies. *Anal Chem.* 2024;96:10935–42.
- Schmid R, Petras D, Nothias LF, Wang M, Aron AT, Jagels A, et al. Ion identity molecular networking for mass spectrometry-based metabolomics in the GNPS environment. *Nat Commun.* 2021;12:3832.
- Wang L, Xing X, Chen L, Yang L, Su X, Rabitz H, et al. Peak Annotation and Verification Engine for Untargeted LC-MS Metabolomics. *Anal Chem.* 2019;91:1838–46.
- Xu YF, Lu W, Rabinowitz JD. Avoiding Misannotation of In-Source Fragmentation Products as Cellular Metabolites in Liquid Chromatography–Mass Spectrometry-Based Metabolomics. *Anal Chem.* 2015;87:2273–81.
- Yu T, Chen JM, Liu W, Zhao JQ, Li P, Liu FJ, et al. In-depth characterization of cycloartane triterpenoids and discovery of species-specific markers from three *Cimicifuga* species guided by a strategy that integrates in-source fragment elimination, diagnostic ion recognition, and feature-based molecular networking. *J Chromatogr A.* 2024;1728:465015.
- El Abied Y, Rutz A, Zuffa S, Amer B, Xing S, Brungs C, et al. Discovery of metabolites prevails amid in-source fragmentation. *Nat Metab.* 2025;7:435–7.
- Pridham TG, Anderson P, Foley C, Lindenfelser LA, Hesseltine CW, Benedict RG. A selection of media for maintenance and taxonomic study of *Streptomyces*. *Antibiot Annu.* 1956;1957:947–53.
- Yang JY, Phelan VV, Simkovsky R, Watrous JD, Trial RM, Fleming TC, et al. Primer on Agar-Based Microbial Imaging Mass Spectrometry. *J Bacteriol.* 2012;194:6023–8.
- Liu J, Zhu X, Kim SJ, Zhang W. Antimycin-type depsipeptides: discovery, biosynthesis, chemical synthesis, and bioactivities. *Nat Prod Rep.* 2016;33:1146–65.
- Schmid R, Heuckeroth S, Korf A, Smirnov A, Myers O, Dyrlund TS, et al. Integrative analysis of multimodal mass spectrometry data in MZmine 3. *Nat Biotechnol.* 2023;41:447–9.
- Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, et al. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods.* 2015;12:523–6.
- Bittremieux W, Avalon NE, Thomas SP, Kakhkhorov SA, Aksenov AA, Gomes PWP, et al. Open access repository-scale propagated nearest neighbor suspect spectral library for untargeted metabolomics. *Nat Commun.* 2023;14:8488.
- Scheubert K, Hufsky F, Petras D, Wang M, Nothias LF, Dührkop K, et al. Significance estimation for large scale metabolomics annotations by spectral matching. *Nat Commun.* 2017;8:1494.
- Seipke RF, Barke J, Brearley C, Hill L, Yu DW, Goss RJM, et al. A Single *Streptomyces* Symbiont Makes Multiple Antifungals to Support the Fungus Farming Ant *Acromyrmex octospinosus*. *PLOS ONE.* 2011;6:e22028.

44. Joynt R, Seipke RFY. A phylogenetic and evolutionary analysis of antimycin biosynthesis. *Microbiology*. 2018;164:28–39.

45. Thomas D, Morris M, Curtis JM, Boyd RK. Fragmentation mechanisms of protonated actinomycins and their use in structural determination of unknown analogues. *J Mass Spectrom*. 1995;30:1111–25.

46. Wang D, Wang C, Gui P, Liu H, Khalaf SMH, Elsayed EA, et al. Identification, Bioactivity, and Productivity of Actinomycins from the Marine-Derived *Streptomyces* heliomycini. *Front Microbiol*. 2017;8: <https://www.frontiersin.org/articles/10.3389/fmicb.2017.01147>

47. Sharma M, Manhas RK. Purification and characterization of actinomycins from *Streptomyces* strain M7 active against methicillin resistant *Staphylococcus aureus* and vancomycin resistant *Enterococcus*. *BMC Microbiol*. 2019;19:44.

48. Machushynets NV, Elsayed SS, Du C, Siegler MA, de la Cruz M, Genilloud O, et al. Discovery of actinomycin L, a new member of the actinomycin family of antibiotics. *Sci Rep*. 2022;12:2813.

49. Yang Z, Qiao Y, Konakalla NC, Strøbech E, Harris P, Peschel G, et al. Streptomyces alleviate abiotic stress in plant by producing pteridic acids. *Nat Commun*. 2023;14:7398.

50. Garcia-Gonzalez E, Müller S, Ensle P, Süßmuth RD, Genersch E. Elucidation of sevadincin, a novel non-ribosomal peptide secondary metabolite produced by the honey bee pathogenic bacterium *Paenibacillus* larvae. *Environ Microbiol*. 2014;16: 1297–309.

51. Gani OABSM, Engh RA. Protein kinase inhibition of clinically important staurosporine analogues. *Nat Prod Rep*. 2010;27:489–98.

52. Stepczynska A, Lauber K, Engels IH, Janssen O, Kabelitz D, Wesselborg S, et al. Staurosporine and conventional anticancer drugs induce overlapping, yet distinct pathways of apoptosis and caspase activation. *Oncogene*. 2001;20:1193–202.

53. Funayama S, Okada K, Iwasaki K, Komiya K, Umezawa I. Structures of trienomycins A, B and C, novel cytoidal ansamycin antibiotics. *J Antibiot*. 1985;38:1677–83.

54. He QR, Tang JJ, Liu Y, Chen ZF, Liu YX, Chen H, et al. The natural product trienomycin A is a STAT3 pathway inhibitor that exhibits potent *in vitro* and *in vivo* efficacy against pancreatic cancer. *Br J Pharm*. 2021;178:2496–515.

55. Nguyen HC, Karray F, Lautru S, Gagnat J, Lebrihi A, Ho Huynh TD, et al. Glycosylation Steps during Spiramycin Biosynthesis in *Streptomyces* ambofaciens: Involvement of Three Glycosyltransferases and Their Interplay with Two Auxiliary Proteins. *Antimicrob Agents Chemother*. 2010;54:2830–9. July.

56. Sugamata R, Sugawara A, Nagao T, Suzuki K, Hirose T, Yamamoto Kichi, et al. Leucomycin A3, a 16-membered macrolide antibiotic, inhibits influenza A virus infection and disease progression. *J Antibiot*. 2014;67:213–22. Mar.

57. Vasanthakumar A, Kattusamy K, Prasad R. Regulation of daunorubicin biosynthesis in *Streptomyces peucetius* – feed forward and feedback transcriptional control. *J Basic Microbiol*. 2013;53:636–44.

58. Kino T, Hatanaka H, Hashimoto M, Nishiyama M, Goto T, Okuhara M, et al. FK-506, a novel immunosuppressant isolated from a *Streptomyces* I. Fermentation, isolation, and physico-chemical and biological characteristics. *J Antibiot*. 1987;40:1249–55.

59. Pires SDS, Oliveira R, Moradas-Ferreira P, V. Mendes M. The Onset of Tacrolimus Biosynthesis in *Streptomyces tsukubaensis* Is Dependent on the Intracellular Redox Status. *Antibiotics*. 2020;9:703.

60. Barona-Gómez F, Wong U, Giannakopoulos AE, Derrick PJ, Challis GL. Identification of a Cluster of Genes that Directs Desferrioxamine Biosynthesis in *Streptomyces coelicolor* M145. *J Am Chem Soc*. 2004;126:16282–3.

61. Ronan JL, Kadi N, McMahon SA, Naismith JH, Alkhafaf LM, Challis GL. Desferrioxamine biosynthesis: diverse hydroxamate assembly by substrate-tolerant acyl transferase DesC. *Philos Trans R Soc B Biol Sci*. 2018;373:20170068.

62. da Silva RR, Wang M, Nothias LF, van der Hooft JJJ, Caraballo-Rodríguez AM, Fox E, et al. Propagating annotations of molecular networks using *in silico* fragmentation. *PLoS Comput Biol*. 2018;14:e1006089.

63. Borelli TC, Arini GS, Feitosa LGP, Dorrestein PC, Lopes NP, da Silva RR. Improving annotation propagation on molecular networks through random walks: introducing ChemWalker. *Bioinformatics*. 2023;39:btad078.

64. Watrous J, Roach P, Alexandrov T, Heath BS, Yang JY, Kersten RD, et al. Mass spectral molecular networking of living microbial colonies. *Proc Natl Acad Sci U A*. 2012;109:E1743–52.

65. Kim HW, Wang M, Leber CA, Nothias LF, Reher R, Kang KB, et al. NPClassifier: A Deep Neural Network-Based Structural Classification Tool for Natural Products. *J Nat Prod*. 2021;84:2795–807.

66. Rammali S, Rahim A, El Aalaoui M, Bencharki B, Dari K, Habach A, et al. Antimicrobial potential of *Streptomyces coeruleofuscus* SCJ isolated from microbiologically unexplored garden soil in Northwest Morocco. *Sci Rep*. 2024;14:3359.

67. Zineb A, Ali Z, Ayoub E, Nadia Z, Assia M, Mohammed M. Optimization of culture conditions for growth and secondary metabolites production by *Streptomyces* sp. MFB27 using response surface methodology. *Microbe*. 2025;8:100423.

68. Vind K, Brunati C, Simone M, Sosio M, Donadio S, Iorio M. Megalochelin, a Tridecapeptide Siderophore from a Talented Streptomycete. *ACS Chem Biol*. 2023. <https://doi.org/10.1021/acschembio.2c00958>

69. Shahnehr MRZ, Strobel M, Vitale GA, Geibel C, Abiead YE, Garg N, et al. ModiFinder: Tandem Mass Spectral Alignment Enables Structural Modification Site Localization. *J Am Soc Mass Spectrom*. 2024;35:2564–78.

70. Belknap KC, Park CJ, Barth BM, Andam CP. Genome mining of biosynthetic and chemotherapeutic gene clusters in *Streptomyces* bacteria. *Sci Rep*. 2020;10:2003.

71. Doroghazi JR, Albright JC, Goering AW, Ju KS, Haines RR, Tchalukov KA, et al. A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat Chem Biol*. 2014;10:963–8.

72. Singh TA, Passari AK, Jajoo A, Bhasin S, Gupta VK, Hashem A, et al. Tapping Into Actinobacterial Genomes for Natural Product Discovery. *Front Microbiol*. 2021. <https://www.frontiersin.org/articles/10.3389/fmicb.2021.655620>

73. Booth TJ, Bozhiyuk KAJ, Liston JD, Batey SFD, Lacey E, Wilkinson B. Bifurcation drives the evolution of assembly-line biosynthesis. *Nat Commun*. 2022;13:3498.

74. Hoffmann MA, Nothias LF, Ludwig M, Fleischauer M, Gentry EC, Witting M, et al. High-confidence structural annotation of metabolites absent from spectral libraries. *Nat Biotechnol*. 2022;40:411–21.

75. Mohimani H, Gurevich A, Mikheenko A, Garg N, Nothias LF, Ninomiya A, et al. Dereplication of peptidic natural products through database search of mass spectra. *Nat Chem Biol*. 2017;13:30–7.

76. van Bergeijk DA, Elsayed SS, Du C, Santiago IN, Roseboom AM, Zhang L, et al. The ubiquitous catechol moiety elicits siderophore and angucycline production in *Streptomyces*. *Commun Chem*. 2022;5:1–12.

77. Caraballo-Rodríguez AM, Dorrestein PC, Pupo MT. Molecular inter-kingdom interactions of endophytes isolated from *Lychنophora ericooides*. *Sci Rep*. 2017;7:5373.

78. Zuffa S, Schmid R, Bauermeister A, Gomes P, Caraballo-Rodríguez PW, El AM, Abiead Y, et al. microbeMASST: a taxonomically informed mass spectrometry search tool for microbial metabolomics data. *Nat Microbiol*. 2024;9:336–45.

79. Wang M, Jarmusch AK, Vargas F, Aksенов AA, Gauglitz JM, Weldon K, et al. Mass spectrometry searches using MASST. *Nat Biotechnol.* 2020;38:23–6.
80. El Abiead Y, Strobel M, Payne T, Fahy E, O'Donovan C, Subramamiam S, et al. Enabling pan-repository reanalysis for big data science of public metabolomics data. *Nat Commun.* 2025;16:4838.
81. Jarmusch AK, Wang M, Aceves CM, Advani RS, Aguirre S, Aksenov AA, et al. ReDU: a framework to find and reanalyze public mass spectrometry data. *Nat Methods.* 2020;17:901–4.
82. Mullowney MW, Duncan KR, Elsayed SS, Garg N, van der Hooft JJ, Martin NI, et al. Artificial intelligence for natural product drug discovery. *Nat Rev Drug Discov.* 2023;11:1–22.
83. Zuo Z, Cao L, Nothia LF, Mohimani H. MS2Planner: improved fragmentation spectra coverage in untargeted mass spectrometry by iterative optimized data acquisition. *Bioinformatics.* 2021;37:i231–6.
84. Banerjee P, Erehman J, Gohlke BO, Wilhelm T, Preissner R, Dunkel M. Super Natural II-a database of natural products. *Nucleic Acids Res.* 2015;43:D935–9.
85. de Matos P, Alcántara R, Dekker A, Ennis M, Hastings J, Haug K, et al. Chemical Entities of Biological Interest: an update. *Nucleic Acids Res.* 2010;38:D249–54.