

Review

Coupling Mass Spectral and Genomic Information to Improve Bacterial Natural Product Discovery Workflows

Max Crüsemann 

Institute for Pharmaceutical Biology, University of Bonn, Nussallee 6, 53115 Bonn, Germany; crusemann@uni-bonn.de

Abstract: Bacterial natural products possess potent bioactivities and high structural diversity and are typically encoded in biosynthetic gene clusters. Traditional natural product discovery approaches rely on UV- and bioassay-guided fractionation and are limited in terms of dereplication. Recent advances in mass spectrometry, sequencing and bioinformatics have led to large-scale accumulation of genomic and mass spectral data that is increasingly used for signature-based or correlation-based mass spectrometry genome mining approaches that enable rapid linking of metabolomic and genomic information to accelerate and rationalize natural product discovery. In this mini-review, these approaches are presented, and discovery examples provided. Finally, future opportunities and challenges for paired omics-based natural products discovery workflows are discussed.

Keywords: bacterial natural products; mass spectrometry; genome mining; paired omics



Citation: Crüsemann, M. Coupling Mass Spectral and Genomic Information to Improve Bacterial Natural Product Discovery Workflows. *Mar. Drugs* **2021**, *19*, 142. <https://doi.org/10.3390/md19030142>

Academic Editor: Vassilios Roussis

Received: 26 January 2021

Accepted: 4 March 2021

Published: 5 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Due to their impressive structural diversity and their wide range of bioactivities, natural products (NP) have been, and are still extensively used by humankind as important sources for drugs [1]. NP structures are generated by the concerted action of biosynthetic enzymes. These are encoded in genes which are, in bacteria, usually grouped to biosynthetic gene clusters (BGCs). This circumstance has facilitated bioinformatics analyses and predictions about the number and classes of natural products that can be synthesized by a bacterial strain [2]. This procedure is termed “genome mining”, a rapidly growing field that has advanced NP research in the last 15 years [3,4]. Sets of closely related BGCs with similar gene content can be grouped into gene cluster families (GCF), that encode the production of identical or highly similar molecules. The recent advances in DNA sequencing have led to a massive accumulation of sequence data in the databases which, in turn, fueled the development of large-scale BGC and GCF analysis pipelines and databases such as BiG-SCAPE [5], BiG-SLICE [6] and BiG-FAM [7] by Medema and colleagues. These frameworks enable the systematic estimation and comparison of NP biosynthetic potential in increasing numbers of bacterial strains.

Biosynthetic machineries usually do not only lead to one single natural product, but may produce a suite of structurally related metabolites through relaxed substrate specificities, causing enzymatic processing of structurally different precursors and intermediates. Mass spectrometry (MS)-based workflows offer opportunities to chart the metabolic diversity that is present in a complex sample, e.g., a crude bacterial extract. The metabolic diversity in complex NP mixtures can be regarded as a collection of “molecular families”, a term for structurally related compounds with related MS fragmentation (MS/MS) spectra [8]. As an outstanding example, the public community data repository and analysis platform GNPS [9,10], developed by Dorrestein, Bandeira and colleagues, offers opportunities for the detailed analysis and visualization of natural product MS/MS fragmentation data by molecular networking. The GNPS environment has also integrated several useful annotation, classification and dereplication tools [11–17] that, if used altogether, aid in

obtaining the maximum amount of information from an MS/MS spectrum or dataset of interest.

One of the most important goals in natural product discovery and the basis for any state-of-the-art biosynthetic study is the direct linkage of a metabolite to its BGC. The classical and most reliable way to establish such a link is either the heterologous expression or the activation of a cryptic BGC, with subsequent detection and characterization of the target compounds in the heterologous or engineered host, or the deletion of the BGC or key biosynthetic genes thereof in the NP producer to abolish production of the natural product of interest. However, although significant advances have been made in these areas [18,19], these approaches are still relatively laborious and time consuming because only a single biosynthetic pathway can be targeted in one experimental workflow, that typically requires several, sometimes cumbersome cloning and transformation procedures. In contrast, MS-guided genome mining techniques that have been developed, enable the parallel establishment of multiple compound-BGC linkages and dereplication in a much more time-effective workflow. The acquisition and in-depth analysis of paired datasets comprising MS/MS data of culture extracts and genome sequences of their producers is thus of promise to accelerate and improve any bacterial NP discovery program.

Concepts and successful examples of linking chemical and biosynthetic space have lately been reviewed by Duncan and coworkers [20]. Another detailed application-oriented review by van der Hoof, Medema and colleagues mainly focused on the key technologies that enable making these linkages [21]. This minireview particularly intends to highlight concepts to directly link mass spectral information and BGCs by (i) signature-based approaches (peptidogenomics and glycogenomics), as well as (ii) correlation-based approaches (pattern-based genome mining, metabologenomics) and provides discovery examples. Finally, latest developments and future promises and challenges in linking biosynthetic and metabolomic data of natural products are presented and discussed.

2. Concepts and Examples for Linking Genomic and Metabolomic Data

2.1. Experiment-Guided Genome Mining: Peptidogenomics and Glycogenomics

These two MS-guided genome mining approaches were developed and pioneered by Kersten, Dorrestein and Moore [22,23]. Both workflows are dependent on the presence of specific signatures, i.e., mass shifts or fragmentation ions, in an MS/MS spectrum from bacterial compound mixtures. These distinctive signatures may be linked to a BGC predicted to encode the biosynthetic machinery to produce NPs with structural motifs to yield these MS/MS fragments. This procedure is particularly applicable for peptides and glycosylated molecules (Figure 1).

In peptidogenomics, the mass shifts relate to the fragmentation of peptides into their constituents, i.e., proteinogenic or modified, non-proteinogenic amino acids, a process that allows for automation [22,24]. A number of subsequent amino acid MS/MS mass shifts constitutes a “sequence tag”, which is instrumental in the search for the respective BGC in the producers’ genome (Figure 1A). For ribosomally synthesized and modified peptide natural products (RiPP), the sequence tag is part of a small, encoded protein, usually clustered with genes encoding posttranslationally modifying enzymes, that is queried in the producers’ genome e.g., by six-frame translations. Nonribosomal peptides (NRP) are synthesized by multimodular assembly line megaenzymes. Here, a detected sequence tag relates to a BGC encoding a sequence of modules with predictable adenylation domain specificities. In glycogenomics, diagnostic mass shifts or fragments are caused by the fragmentation of bonds to sugars or, preferably, modified deoxysugars, both frequently observed features of bioactive natural products [23]. The biosynthesis of deoxysugars is typically encoded by subclusters of modifying biosynthetic genes and glycosyltransferases, clustered with genes encoding core NP biosynthetic machineries (e.g., polyketide synthases) and can be matched with the detected MS/MS deoxysugar fragment(s) (Figure 1B).

In a landmark study in 2011, the concept of peptidogenomics was introduced to the NP community and systematically used to uncover and characterize a series of novel RiPPs

from well investigated *Streptomyces* strains such as *S. lividans*, *S. coelicolor* and *S. griseus*. Additionally, five novel analogs of the nonribosomal lipopeptide stendomycin were characterized from *S. hygroscopicus* and connected to their BGC (Figure 1A) [22]. In a subsequent study, *Streptomyces roseosporus* natural products were mapped with a combination of molecular networking and peptidogenomics which led to the discovery of the stenothricin BGC [25]. Another peptidogenomic study on *S. roseosporus* based on imaging MS revealed that the potent antibiotic peptide arylomycin was of nonribosomal origin [26]. Bromoalterochromides were discovered and connected to their BGC in two marine *Pseudoalteromonas* bacteria from a large scale nano-DESI MS/MS dataset of *Bacillus* and *Pseudoalteromonas* strains [8]. From the plant pathogen *Ralstonia solanacearum*, the bioactive lipopeptide ralsolamycin was also identified using the peptidogenomic approach [27].

The peptidogenomic concept was automated by Pevzner, Mohimani and colleagues, leading to the development of automated peptidogenomics tools specifically designed for RiPPs (RiPP-Quest) [28], NRPs (NRP-Quest) [29] and both (Pep2Path) [30]. Application of RiPP-Quest enabled the discovery of informatipeptin (Figure 1C), a new class III lanthipeptide from *Streptomyces viridochromogenes*. Recently, the development of MetaMiner, an advancement of the RiPP-Quest tool, designed for the query of larger datasets e.g., from metagenomes, and its application to several datasets in the GNPS database led to the discovery and annotation of seven previously unknown RiPPs [31].

The development of the glycogenomic approach and its proof-of-principle application to a set of marine actinomycete crude extracts enabled discovery and MS-guided isolation of arenimycin B (Figure 1D), a type II polyketide comprising the two characteristic deoxysugar moieties forosamine and O-methyl rhamnose from the marine actinomycete *Salinispora arenicola* CNB527 [23]. The derivative arenimycin A, containing only O-methyl rhamnose, had previously been isolated from another *S. arenicola* strain [32] by classical, UV-guided purification. However, analysis of the biosynthetic potential of CNB527 suggested that a candidate BGC for arenimycin biosynthetic machinery could be capable of adding another sugar moiety to the molecule. After its characterization, it was thus concluded that arenimycin B is actually the end product of the biosynthetic pathway, and was notably found to be more bioactive than the previously isolated arenimycin A, showing a twofold or greater increase in activity against clinically relevant, multidrug-resistant strains of *Staphylococcus aureus* [23]. Another glycogenomic example with marine origin is the discovery of five rosamicin derivatives from the marine actinomycete *Salinispora pacifica* CNS237 (Figure 1B) [33]. This group of antibioticly active, glycosylated polyketides, among them three unprecedented analogs, was discovered by their characteristic desosamine fragment from a large MS/MS dataset that was used to prioritize marine actinomycete strains by molecular networking [34]. After their dereplication and subsequent structure elucidation, it was later revealed that these compounds are actually the end product of their polyketide synthase (PKS) assembly line pathway that is, however, also responsible for the production of salinipyronone and pacificanone, linear polyketides that had previously been isolated by a classical approach [35]. Unexpectedly, both appeared to be shunt products of the PKS, as proven by mutagenesis experiments in the rosamicin assembly line [33]. Analogously to peptidogenomics, glycogenomics also holds potential for automation, although this has not yet been implemented.

2.2. Correlation-Based Approaches on Larger Paired Datasets: Pattern-Based Genome Mining, Metabologenomics

Another possibility to link genomic with metabolomic information is the application of correlation-based approaches. Here metabolite patterns, obtained from larger MS datasets of sequenced bacteria, are compared and correlated with their BGC or GCF patterns, derived from comparative analyses of a set of genomes (Figure 2). Notably, these correlations are independent of the chemical class of the detected metabolites. Talented NP producers such as actinomycetes harbor a multitude of BGCs, whereas taxonomically closely related strains characteristically possess overlapping patterns of encoded BGCs. This means that homologous BGCs are frequently encoded in more than one or several

related strains, while other BGCs are unique for particular strains. An illustrative model for this phenomenon is the marine actinomycete genus *Salinispora* [36].

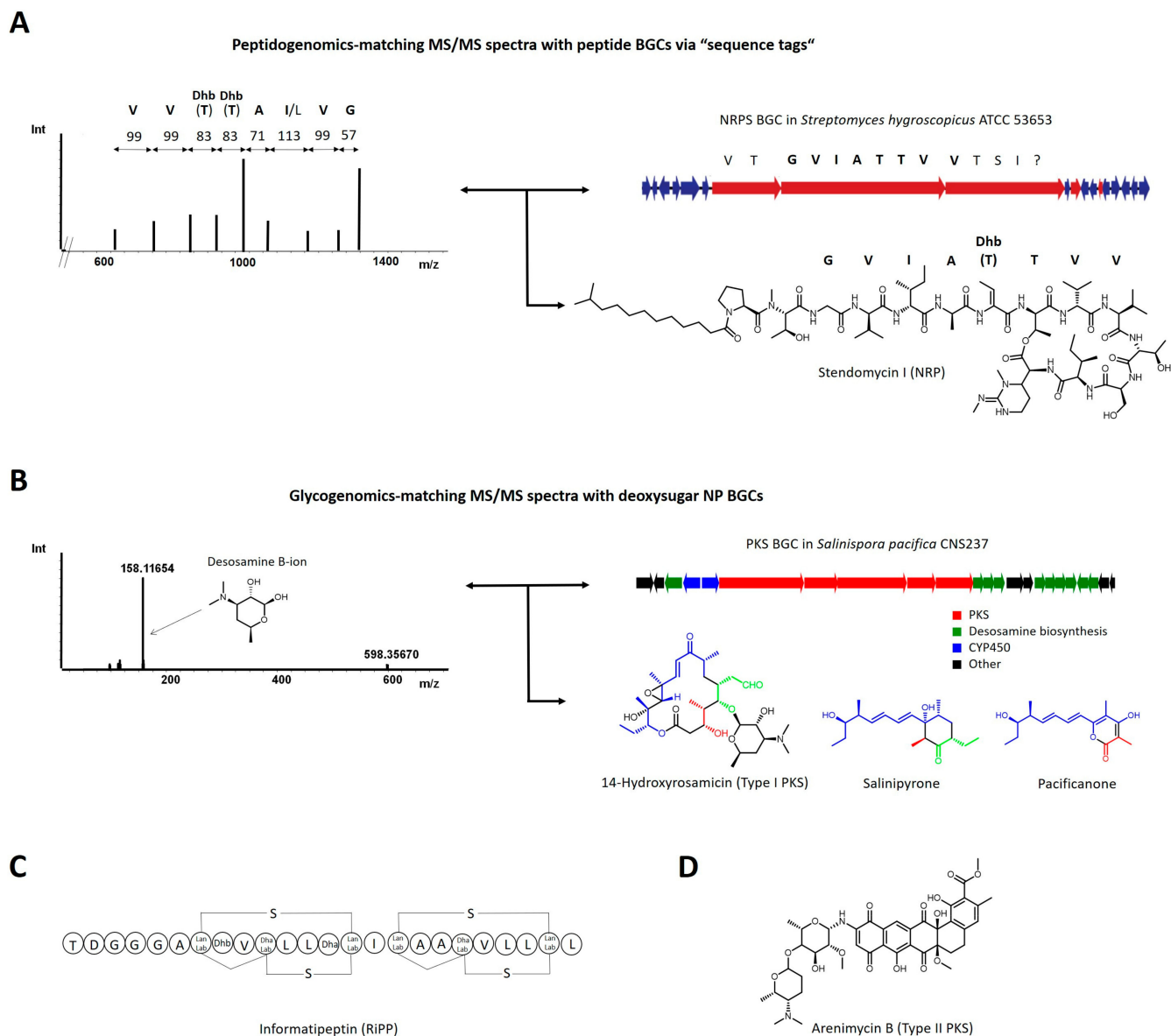


Figure 1. Concepts and discovery examples for experiment-guided genome mining. (A) Peptidogenomics: *Streptomyces hygroscopicus* MS/MS data yielded a sequence tag with eight amino acids, two of them dehydrated threonines (Dhb). The sequence tag matched with a sequence of adenylation domain predictions of an orphan nonribosomal peptide synthetase BGC, facilitating the targeted isolation and characterization of stendomycin lipopeptides [22]. The second threonine dehydration appeared only during MS/MS fragmentation as elimination product of the ester bond. (B) Glycogenomics: Matching of MS/MS spectra of *Salinispora pacifica* CNS237 with a type I polyketide synthase BGC encoding deoxysugar biosynthesis genes revealed several rosamicin derivatives and enabled their targeted isolation and further characterization. The previously isolated linear polyketides salinipyrrone and pacificanone appear to be shunt products of the rosamicin PKS, revealed by mutagenesis experiments. Building blocks synthesized by the same module(s) are color-coded accordingly [33]. (C) Further natural products from different classes discovered by the peptidogenomic [28] and (D) glycogenomic approach [23]. For more details regarding these concepts, please refer to references [22,23].

Salinispora species and strains are very closely related on 16S-RNA level, but can be discriminated by the presence of species- or strain-specific patterns of encoded NP BGCs (compare Figure 2A, left). In a remarkable genome mining study led by Ziemert and Jensen, 75 *Salinispora* strains were analyzed and compared regarding their PKS and NRPS pathway

variety and evolution [37]. This comprehensive analysis was later extended to 119 strains and to other pathway types such as terpenes and RiPPs [38].

In a pioneering study from the Jensen and Moore labs, the metabolomes of 35 *Salinispora* strains were visualized with GNPS molecular networking and then compared with the NRPS/PKS BGC patterns of the respective strains to establish compound-BGC links [39]. These correlative analyses enabled the linkage of an orphan BGC to the polyketide arenicolide and the targeted isolation, structure elucidation and biological characterization of the cytotoxic, echinomycin-like nonribosomal depsipeptide retimycin A, that was encoded and produced by only one strain in the collection, *S. arenicola* CNT005 (Figure 2A).

An analogous study was performed in the bacterial genera *Photorhabdus* and *Xenorhabdus*, both associated with insects and known for prolific natural product production, by the Bode group [40]. Here, a metabolomic network from HPLC-MS/MS data of 30 strains was created, annotated and compared with BGC patterns in the respective strains (Figure 2B). This study revealed the robust expression of known metabolites under laboratory conditions in a number of strains, but also led to the detection of previously unidentified metabolite classes in these bacteria, such as the novel xefoampeptides and tilivallin and connection to their BGCs. Furthermore, novel depsipeptides named fatflabets and xeneprotides were discovered from analysis of the molecular network, and their structures elucidated. However, a complete BGC for these novel compound families could not be assigned with certainty.

A similar concept to bridge metabolomic and genomic information, termed metabologenomics, was developed in the Metcalf and Kelleher labs (Figure 2C). This approach relies on the establishment of correlations between MS spectra and GCFs in a huge dataset of 830 sequenced actinomycete bacteria, of which 178 were subjected to detailed HPLC-MS metabolic profiling in different culture media [41]. Here, a correlation score between GCF and MS1 data was generated and then applied by searching for exact masses of predicted metabolites in the dataset. Subsequent mining of this extensive, paired dataset for detected metabolites encoded by biochemically interesting BGCs enabled the discovery and characterization of several natural products such as tambromycin [42], rimosamides [43] and tyrobetaines [44] and the detailed investigation of their biosynthetic pathways.

Zdouc, Sosio and colleagues recently performed a detailed metabolomic investigation of the actinobacterial genus *Planomonospora* [45]. Four of the 72 investigated strains were also genome-sequenced, which allowed for a paired omics analysis leading to the annotation of a BGC for the thiopeptide siomycin and congeners. Furthermore, two novel biaryl-linked tripeptides were isolated after network analysis and their structures elucidated. They represent the first members of a widespread novel class of small RiPPs, encoded by the smallest gene ever reported, as revealed by peptidogenomics and heterologous expression [46]. Metabologenomics was also used by the Duncan lab for the evaluation of a dataset of 25 polar actinomycetes, published in this Special Issue [47]. Their metabolomes were analyzed and correlated to genome data by using a newly developed tool, NP-linker, designed for the automated establishment of NP-BGC correlations [48].

In a recent study on the biosynthetic and metabolic diversity in the actinomycete genus *Nocardia*, metabolite-BGC correlations were analyzed based on a double-network approach by the Ziemert and Kaysser groups [49]: A metabolomic network was constructed with GNPS molecular networking [9], as well as a BGC network of all selected strains created via BiG-SCAPE [5]. Then, both networks were analyzed and compared for correlations of molecular families and gene cluster families over the same number of strains. This strategy was validated by the strain-specific discovery and annotation of a battery of unprecedented nocobactin-like siderophores.

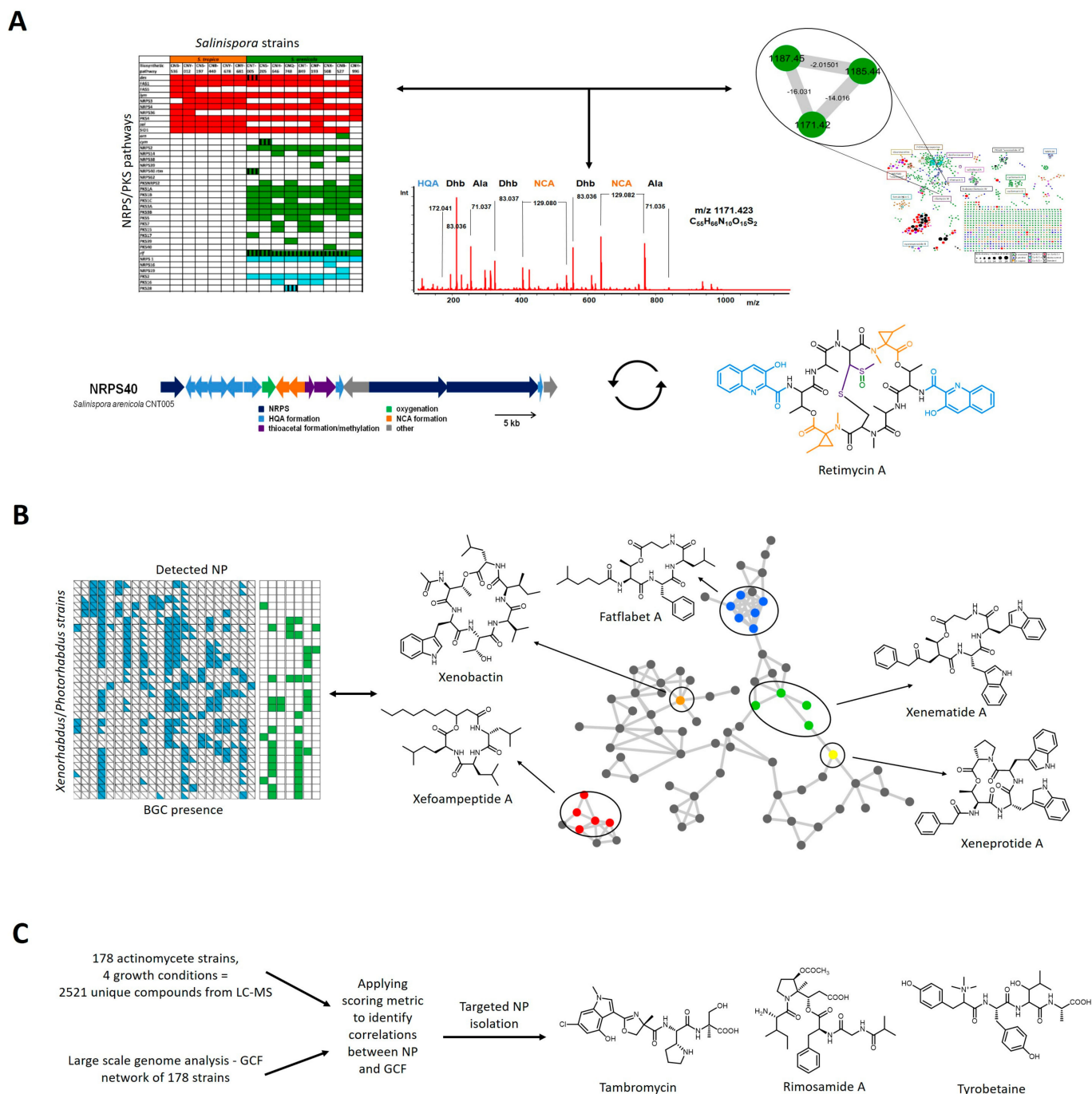


Figure 2. Discovery examples for correlation-based approaches using paired datasets. (A) Correlation of *Salinispora* strain BGC patterns with a molecular network of 35 strains led to the identification of a candidate peptide, encoded and produced by only one strain in the dataset. The peptide was matched to its NRPS BGC with additional help of the peptidogenomic approach. The structure of the elucidated metabolite retimycin A is depicted. Building blocks are color-coded corresponding to responsible biosynthetic genes [39]. Taken from reference 39 and rearranged with permission of Elsevier. **(B)** *Xenorhabdus* and *Photorhabdus* strains were analyzed for BGC patterns and production of encoded metabolites (left). Subsequent molecular network analysis led to the identification and discovery of several NRPS-derived cyclic depsipeptides (right) [40]. **(C)** Metabologenomic workflow of a 178 strain actinomycetes dataset [41–44]. An example for the applied scoring metric can be found in reference 21.

The generation of standardized community repositories such as MIBiG for characterized NP BGCs [50], and the GNPS database for MS/MS data of NP datasets and compounds [9], has improved and facilitated many natural products workflows. However, genome-metabolome links have not been systematically documented and are cumbersome to search for. To overcome this obstacle and to standardize NP-BGC links that can be reused by the community for further projects, recently the paired omics database has been developed and launched [51]. This platform gathers a large number of paired datasets generated by the NP community including links to MS/MS datasets on GNPS and sequence data of characterized BGCs on MIBiG and standardized metabolite-BGC links were generated. This standardized and open community database may be very useful for the application and automation of future correlative network-based approaches and prioritization of novel metabolites and BGCs that are worth investigating.

An alternative pipeline for bioinformatic analyses of BGCs and compound-cluster matching was developed by the Magarvey lab [52]: Here, a prediction engine (PRISM) identifies and predicts BGC in microbial genomes. A retrobiosynthetic algorithm (GRAPE) performs retrobiosynthetic analyses of known natural products and suggests a likely BGC for these metabolites. A matching algorithm (GARLIC) then compares PRISM and GRAPE outputs and gives matching scores. By that procedure, BGCs with unknown products can be identified with high confidence. Additionally, a “genomes to natural products” (GNP) algorithm matches LC-MS/MS data to BGCs by structure prediction, substructure analysis and in silico fragmentation prediction generating confidence scores of the NP-BGC matches [53]. Notably, this pipeline is restricted to modular PKS and NRPS pathways.

To conclude, in the last decade, several novel MS-guided genome mining workflows and global, standardized mass spectral and BGC databases have been developed and led to a significant number of natural product discoveries. For reliable NP-BGC linkages, these paired omics workflows rely on high quality MS/MS and sequence data and bioinformatic, mass spectral and biosynthetic knowledge. To further expand a paired omics NP mining workflow to other natural product classes, the integrated use of recently developed substructure annotation tools [11], classification-based methods [54] and fragmentation trees [55] together with the use of further improved automated linking approaches [48] are of great promise. Structure elucidation remains a major bottleneck in NP discovery pipelines, and is often limited by the low yields of the NP of interest. However, the development of neural network algorithms for NMR analysis [56] and novel structure elucidation methods such as MicroED [57] may, integrated into the described workflows, further brighten the future for natural product research and enable many exciting discoveries.

Funding: The author acknowledges a postdoctoral fellowship from the Deutsche Forschungsgemeinschaft (DFG), grant number CR464-1.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Newman, D.J.; Cragg, G.M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **2020**, *83*, 770–803. [[CrossRef](#)] [[PubMed](#)]
2. Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S.Y.; Medema, M.H.; Weber, T. antiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **2019**, *47*, W81–W87. [[PubMed](#)]
3. Ziemert, N.; Alanjary, M.; Weber, T. The evolution of genome mining in microbes—A review. *Nat. Prod. Rep.* **2016**, *33*, 988–1005. [[CrossRef](#)] [[PubMed](#)]
4. Kenshole, E.; Herisse, M.; Michael, M.; Pidot, S.J. Natural product discovery through microbial genome mining. *Curr. Opin. Chem. Biol.* **2020**, *60*, 47–54. [[CrossRef](#)]
5. Navarro-Muñoz, J.C.; Selem-Mojica, N.; Mullowney, M.W.; Kautsar, S.A.; Tryon, J.H.; Parkinson, E.I.; De Los Santos, E.L.C.; Yeong, M.; Cruz-Morales, P.; Abubucker, S.; et al. A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* **2020**, *16*, 60–68. [[CrossRef](#)] [[PubMed](#)]
6. Kautsar, S.A.; Van der Hooft, J.J.J.; De Ridder, D.; Medema, M.H. BiG-SLiCE: A highly scalable tool maps the diversity of 1.2 million biosynthetic gene clusters. *Gigascience* **2021**, *10*, giaa154. [[CrossRef](#)] [[PubMed](#)]

7. Kautsar, S.A.; Blin, K.; Shaw, S.; Weber, T.; Medema, M.H. BiG-FAM: The biosynthetic gene cluster families database. *Nucleic Acids Res.* **2021**, *49*, D490–D497. [[CrossRef](#)]
8. Nguyen, D.D.; Wu, C.H.; Moree, W.J.; Lamsa, A.; Medema, M.H.; Zhao, X.; Gavilan, R.G.; Aparicio, M.; Atencio, L.; Jackson, C.; et al. MS/MS networking guided analysis of molecule and gene cluster families. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 2611–2620. [[CrossRef](#)] [[PubMed](#)]
9. Wang, M.; Carver, J.J.; Phelan, V.V.; Sanchez, L.M.; Garg, N.; Peng, Y.; Nguyen, D.D.; Watrous, J.; Kaponov, C.A.; Luzzatto-Knaan, T.; et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34*, 828–837. [[CrossRef](#)] [[PubMed](#)]
10. Aron, A.T.; Gentry, E.C.; McPhail, K.L.; Nothias, L.F.; Nothias-Esposito, M.; Bouslimani, A.; Petras, D.; Gauglitz, J.M.; Sikora, N.; Vargas, F.; et al. Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nat. Protoc.* **2020**, *15*, 1954–1991. [[CrossRef](#)]
11. Van der Hooft, J.J.J.; Wandy, J.; Young, F.; Padmanabhan, S.; Gerasimidis, K.; Burgess, K.E.V.; Barrett, M.P.; Rogers, S. Unsupervised Discovery and Comparison of Structural Families Across Multiple Samples in Untargeted Metabolomics. *Anal. Chem.* **2017**, *89*, 7569–7577. [[CrossRef](#)] [[PubMed](#)]
12. Da Silva, R.R.; Wang, M.; Nothias, L.F.; Van der Hooft, J.J.J.; Caraballo-Rodríguez, A.M.; Fox, E.; Balunas, M.J.; Klassen, J.L.; Lopes, N.P.; Dorrestein, P.C. Propagating annotations of molecular networks using in silico fragmentation. *PLoS Comput. Biol.* **2018**, *14*, e1006089. [[CrossRef](#)] [[PubMed](#)]
13. Ernst, M.; Kang, K.B.; Caraballo-Rodríguez, A.M.; Nothias, L.F.; Wandy, J.; Chen, C.; Wang, M.; Rogers, S.; Medema, M.H.; Dorrestein, P.C.; et al. MolNetEnhancer: Enhanced Molecular Networks by Integrating Metabolome Mining and Annotation Tools. *Metabolites* **2019**, *9*, 144. [[CrossRef](#)]
14. Jarmusch, A.K.; Wang, M.; Aceves, C.M.; Advani, R.S.; Aguirre, S.; Aksenov, A.A.; Aleti, G.; Aron, A.T.; Bauermeister, A.; Bolleddu, S.; et al. ReDU: A framework to find and reanalyze public mass spectrometry data. *Nat. Methods* **2020**, *17*, 901–904. [[CrossRef](#)] [[PubMed](#)]
15. Wang, M.; Jarmusch, A.K.; Vargas, F.; Aksenov, A.A.; Gauglitz, J.M.; Weldon, K.; Petras, D.; Da Silva, R.R.; Quinn, R.A.; Melnik, A.V.; et al. Mass spectrometry searches using MASST. *Nat. Biotechnol.* **2020**, *38*, 23–26. [[CrossRef](#)]
16. Mohimani, H.; Gurevich, A.; Mikheenko, A.; Garg, N.; Nothias, L.F.; Ninomiya, A.; Takada, K.; Dorrestein, P.C.; Pevzner, P.A. Dereplication of peptidic natural products through database search of mass spectra. *Nat. Chem. Biol.* **2017**, *13*, 30–37. [[CrossRef](#)]
17. Mohimani, H.; Gurevich, A.; Shlemov, A.; Mikheenko, A.; Korobeynikov, A.; Cao, L.; Shcherbin, E.; Nothias, L.F.; Dorrestein, P.C.; Pevzner, P.A. Dereplication of microbial metabolites through database search of mass spectra. *Nat. Commun.* **2018**, *9*, 4035. [[CrossRef](#)]
18. Zhang, J.J.; Tang, X.; Moore, B.S. Genetic platforms for heterologous expression of microbial natural products. *Nat. Prod. Rep.* **2019**, *36*, 1313–1332. [[CrossRef](#)] [[PubMed](#)]
19. Tong, Y.; Weber, T.; Lee, S.Y. CRISPR/Cas-based genome engineering in natural product discovery. *Nat. Prod. Rep.* **2019**, *36*, 1262–1280. [[CrossRef](#)] [[PubMed](#)]
20. Soldatou, S.; Eldjarn, G.H.; Huerta-Urbe, A.; Rogers, S.; Duncan, K.R. Linking biosynthetic and chemical space to accelerate microbial secondary metabolite discovery. *FEMS Microbiol. Lett.* **2019**, *366*, fnz142. [[CrossRef](#)]
21. Van der Hooft, J.J.J.; Mohimani, H.; Bauermeister, A.; Dorrestein, P.C.; Duncan, K.R.; Medema, M.H. Linking genomics and metabolomics to chart specialized metabolic diversity. *Chem Soc. Rev.* **2020**, *49*, 3297–3314. [[CrossRef](#)]
22. Kersten, R.D.; Yang, Y.L.; Xu, Y.; Cimermancic, P.; Nam, S.J.; Fenical, W.; Fischbach, M.A.; Moore, B.S.; Dorrestein, P.C. A mass spectrometry-guided genome mining approach for natural product peptidogenomics. *Nat. Chem. Biol.* **2011**, *7*, 794–802. [[CrossRef](#)]
23. Kersten, R.D.; Ziemert, N.; Gonzalez, D.J.; Duggan, B.M.; Nizet, V.; Dorrestein, P.C.; Moore, B.S. Glycogenomics as a mass spectrometry-guided genome-mining method for microbial glycosylated molecules. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, E4407–E4416. [[CrossRef](#)] [[PubMed](#)]
24. Mohimani, H.; Pevzner, P.A. Dereplication, sequencing and identification of peptidic natural products: From genome mining to peptidogenomics to spectral networks. *Nat. Prod. Rep.* **2016**, *33*, 73–86. [[CrossRef](#)]
25. Liu, W.T.; Lamsa, A.; Wong, W.R.; Boudreau, P.D.; Kersten, R.D.; Peng, Y.; Moree, W.J.; Duggan, B.M.; Moore, B.S.; Gerwick, W.H.; et al. MS/MS-based networking and peptidogenomics guided genome mining revealed the stenothricin gene cluster in *Streptomyces roseosporus*. *J. Antibiot.* **2014**, *67*, 99–104. [[CrossRef](#)] [[PubMed](#)]
26. Liu, W.T.; Kersten, R.D.; Yang, Y.L.; Moore, B.S.; Dorrestein, P.C. Imaging mass spectrometry and genome mining via short sequence tagging identified the anti-infective agent arylomycin in *Streptomyces roseosporus*. *J. Am. Chem. Soc.* **2011**, *133*, 18010–18013. [[CrossRef](#)]
27. Spraker, J.E.; Sanchez, L.M.; Lowe, T.M.; Dorrestein, P.C.; Keller, N.P. *Ralstonia solanacearum* lipopeptide induces chlamydospore development in fungi and facilitates bacterial entry into fungal tissues. *ISME J.* **2016**, *10*, 2317–2330. [[CrossRef](#)] [[PubMed](#)]
28. Mohimani, H.; Kersten, R.D.; Liu, W.T.; Wang, M.; Purvine, S.O.; Wu, S.; Brewer, H.M.; Pasa-Tolic, L.; Bandeira, N.; Moore, B.S.; et al. Automated genome mining of ribosomal peptide natural products. *ACS Chem. Biol.* **2014**, *9*, 1545–1551. [[CrossRef](#)]
29. Mohimani, H.; Liu, W.T.; Kersten, R.D.; Moore, B.S.; Dorrestein, P.C.; Pevzner, P.A. NRPquest: Coupling Mass Spectrometry and Genome Mining for Nonribosomal Peptide Discovery. *J. Nat. Prod.* **2014**, *77*, 1902–1909. [[CrossRef](#)] [[PubMed](#)]
30. Medema, M.H.; Paalvast, Y.; Nguyen, D.D.; Melnik, A.; Dorrestein, P.C.; Takano, E.; Breitling, R. Pep2Path: Automated mass spectrometry-guided genome mining of peptidic natural products. *PLoS Comput. Biol.* **2014**, *10*, e1003822. [[CrossRef](#)]

31. Cao, L.; Gurevich, A.; Alexander, K.L.; Naman, C.B.; Leão, T.; Glukhov, E.; Luzzatto-Knaan, T.; Vargas, F.; Quinn, R.A.; Bouslimani, A.; et al. MetaMiner: A Scalable Peptidogenomics Approach for Discovery of Ribosomal Peptide Natural Products with Blind Modifications from Microbial Communities. *Cell Syst.* **2019**, *9*, 600–608.e4. [[CrossRef](#)]
32. Asolkar, R.N.; Kirkland, T.N.; Jensen, P.R.; Fenical, W. Arenimycin, an antibiotic effective against rifampin- and methicillin-resistant *Staphylococcus aureus* from the marine actinomycete *Salinispora arenicola*. *J. Antibiot.* **2010**, *63*, 37–39. [[CrossRef](#)]
33. Awakawa, T.; Crüseman, M.; Munguia, J.; Ziemert, N.; Nizet, V.; Fenical, W.; Moore, B.S. Salinipyronone and Pacificanone Are Biosynthetic By-products of the Rosamicin Polyketide Synthase. *Chembiochem* **2015**, *16*, 1443–1447. [[CrossRef](#)] [[PubMed](#)]
34. Crüseman, M.; O'Neill, E.C.; Larson, C.B.; Melnik, A.V.; Floros, D.J.; Da Silva, R.R.; Jensen, P.R.; Dorrestein, P.C.; Moore, B.S. Prioritizing Natural Product Diversity in a Collection of 146 Bacterial Strains Based on Growth and Extraction Protocols. *J. Nat. Prod.* **2017**, *80*, 588–597. [[CrossRef](#)]
35. Oh, D.C.; Gontang, E.A.; Kauffman, C.A.; Jensen, P.R.; Fenical, W. Salinipyrones and pacificanones, mixed-precursor polyketides from the marine actinomycete *Salinispora pacifica*. *J. Nat. Prod.* **2008**, *71*, 570–575. [[CrossRef](#)] [[PubMed](#)]
36. Jensen, P.R.; Moore, B.S.; Fenical, W. The marine actinomycete genus *Salinispora*: A model organism for secondary metabolite discovery. *Nat. Prod. Rep.* **2015**, *32*, 738–751. [[CrossRef](#)] [[PubMed](#)]
37. Ziemert, N.; Lechner, A.; Wietz, M.; Millán-Aguiñaga, N.; Chavarria, K.L.; Jensen, P.R. Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E1130–E1139. [[CrossRef](#)] [[PubMed](#)]
38. Letzel, A.C.; Li, J.; Amos, G.C.A.; Millán-Aguiñaga, N.; Ginigini, J.; Abdelmohsen, U.R.; Gaudêncio, S.P.; Ziemert, N.; Moore, B.S.; Jensen, P.R. Genomic insights into specialized metabolism in the marine actinomycete *Salinispora*. *Environ. Microbiol.* **2017**, *19*, 3660–3673. [[CrossRef](#)]
39. Duncan, K.R.; Crüseman, M.; Lechner, A.; Sarkar, A.; Li, J.; Ziemert, N.; Wang, M.; Bandeira, N.; Moore, B.S.; Dorrestein, P.C.; et al. Molecular networking and pattern-based genome mining improves discovery of biosynthetic gene clusters and their products from *Salinispora* species. *Chem. Biol.* **2015**, *22*, 460–471. [[CrossRef](#)] [[PubMed](#)]
40. Tobias, N.J.; Wolff, H.; Djahanschiri, B.; Grundmann, F.; Kronenwerth, M.; Shi, Y.M.; Simonyi, S.; Grün, P.; Shapiro-Ilan, D.; Pidot, S.J.; et al. Natural product diversity associated with the nematode symbionts *Photorhabdus* and *Xenorhabdus*. *Nat. Microbiol.* **2017**, *2*, 1676–1685. [[CrossRef](#)] [[PubMed](#)]
41. Doroghazi, J.R.; Albright, J.C.; Goering, A.W.; Ju, K.S.; Haines, R.R.; Tchalukov, K.A.; Labeda, D.P.; Kelleher, N.L.; Metcalf, W.W. A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat. Chem. Biol.* **2014**, *10*, 963–968. [[CrossRef](#)]
42. Goering, A.W.; McClure, R.A.; Doroghazi, J.R.; Albright, J.C.; Haverland, N.A.; Zhang, Y.; Ju, K.S.; Thomson, R.J.; Metcalf, W.W.; Kelleher, N.L. Metabologenomics: Correlation of microbial gene clusters with metabolites drives discovery of a nonribosomal peptide with an unusual amino acid monomer. *ACS Cent. Sci.* **2016**, *2*, 99–108. [[CrossRef](#)] [[PubMed](#)]
43. McClure, R.A.; Goering, A.W.; Ju, K.S.; Baccile, J.A.; Schroeder, F.C.; Metcalf, W.W.; Thomson, R.J.; Kelleher, N.L. Elucidating the Rimosamide-Detoxin Natural Product Families and Their Biosynthesis Using Metabolite/Gene Cluster Correlations. *ACS Chem. Biol.* **2016**, *11*, 3452–3460. [[CrossRef](#)]
44. Parkinson, E.I.; Tryon, J.H.; Goering, A.W.; Ju, K.S.; McClure, R.A.; Kemball, J.D.; Zhukovsky, S.; Labeda, D.P.; Thomson, R.J.; Kelleher, N.L.; et al. Discovery of the Tyrobetaine Natural Products and Their Biosynthetic Gene Cluster via Metabologenomics. *ACS Chem. Biol.* **2018**, *13*, 1029–1037. [[CrossRef](#)]
45. Zdouc, M.M.; Iorio, M.; Maffioli, S.I.; Crüseman, M.; Donadio, S.; Sosio, M. Planomonospora: A Metabolomics Perspective on an Underexplored Actinobacteria Genus. *J. Nat. Prod.* **2021**, *84*, 204–219. [[CrossRef](#)]
46. Zdouc, M.M.; Alanjary, M.M.; Zarazúa, G.S.; Maffioli, S.I.; Crüseman, M.; Medema, M.H.; Donadio, S.; Sosio, M. A biaryl-linked tripeptide from *Planomonospora* reveals a widespread class of minimal RiPP gene clusters. *Cell Chem. Biol.* **2020**. [[CrossRef](#)]
47. Soldatou, S.; Eldjárn, G.H.; Ramsey, A.; Van der Hooft, J.J.J.; Hughes, A.H.; Rogers, S.; Duncan, K.R. Comparative Metabologenomics Analysis of Polar Actinomycetes. *Marine Drugs* **2021**, *19*, 103. [[CrossRef](#)] [[PubMed](#)]
48. Eldjárn, G.H.; Ramsay, A.; Van der Hooft, J.J.J.; Duncan, K.R.; Soldatou, S.; Rousu, R.; Daly, R.; Wandy, J.; Rogers, S. Ranking microbial metabolomic and genomic links in the NPLinker framework using complementary scoring functions. *bioRxiv* **2020**. [[CrossRef](#)]
49. Männle, D.; McKinnie, S.M.K.; Mantri, S.S.; Steinke, K.; Lu, Z.; Moore, B.S.; Ziemert, N.; Kaysser, L. Comparative Genomics and Metabolomics in the Genus *Nocardia*. *mSystems* **2020**, *5*. [[CrossRef](#)]
50. Kautsar, S.A.; Blin, K.; Shaw, S.; Navarro-Muñoz, J.C.; Terlouw, B.R.; Van der Hooft, J.J.J.; Van Santen, J.A.; Tracanna, V.; Suarez Duran, H.G.; Pascal Andreu, V.; et al. MIBiG 2.0: A repository for biosynthetic gene clusters of known function. *Nucleic Acids Res.* **2020**, *48*, D454–D458. [[CrossRef](#)] [[PubMed](#)]
51. Schorn, M.A.; Verhoeven, S.; Ridder, L.; Huber, F.; Acharya, D.D.; Aksenov, A.A.; Aleti, G.; Amiri Moghaddam, J.; Aron, A.T.; Aziz, S.; et al. A community resource for paired genomic and metabolomic data mining. *Nat. Chem. Biol.* **2021**. [[CrossRef](#)]
52. Dejong, C.A.; Chen, G.M.; Li, H.; Johnston, C.W.; Edwards, M.R.; Rees, P.N.; Skinnider, M.A.; Webster, A.L.; Magarvey, N.A. Polyketide and nonribosomal peptide retro-biosynthesis and global gene cluster matching. *Nat. Chem. Biol.* **2016**, *12*, 1007–1014. [[CrossRef](#)] [[PubMed](#)]
53. Johnston, C.W.; Skinnider, M.A.; Wyatt, M.A.; Li, X.; Ranieri, M.R.; Yang, L.; Zechel, D.L.; Ma, B.; Magarvey, N.A. An automated Genomes-to-Natural Products platform (GNP) for the discovery of modular natural products. *Nat. Commun.* **2015**, *6*, 8421. [[CrossRef](#)] [[PubMed](#)]

-
54. Dührkop, K.; Nothias, L.F.; Fleischauer, M.; Reher, R.; Ludwig, M.; Hoffmann, M.A.; Petras, D.; Gerwick, W.H.; Rousu, J.; Dorrestein, P.C.; et al. Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. *Nat. Biotechnol.* **2020**. [[CrossRef](#)] [[PubMed](#)]
 55. Tripathi, A.; Vázquez-Baeza, Y.; Gauglitz, J.M.; Wang, M.; Dührkop, K.; Nothias-Esposito, M.; Acharya, D.D.; Ernst, M.; Van der Hoof, J.J.J.; Zhu, Q.; et al. Chemically informed analyses of metabolomics mass spectrometry data with Qemistree. *Nat. Chem. Biol.* **2021**, *17*, 146–151. [[CrossRef](#)] [[PubMed](#)]
 56. Reher, R.; Kim, H.W.; Zhang, C.; Mao, H.H.; Wang, M.; Nothias, L.F.; Caraballo-Rodriguez, A.M.; Glukhov, E.; Teke, B.; Leao, T.; et al. A Convolutional Neural Network-Based Approach for the Rapid Annotation of Molecularly Diverse Natural Products. *J. Am. Chem. Soc.* **2020**, *142*, 4114–4120. [[CrossRef](#)] [[PubMed](#)]
 57. Danelius, E.; Halaby, S.; Van der Donk, W.A.; Gonen, T. MicroED in natural product and small molecule research. *Nat. Prod. Rep.* **2020**. [[CrossRef](#)]