



Review Recent Advances in Discovery of Lead Structures from Microbial Natural Products: Genomics- and Metabolomics-Guided Acceleration

Linda Sukmarini 匝

Research Center for Biotechnology, Indonesian Institute of Sciences (LIPI), Cibinong, Bogor 16911, West Java, Indonesia; linda.sukmarini@lipi.go.id

Abstract: Natural products (NPs) are evolutionarily optimized as drug-like molecules and remain the most consistently successful source of drugs and drug leads. They offer major opportunities for finding novel lead structures that are active against a broad spectrum of assay targets, particularly those from secondary metabolites of microbial origin. Due to traditional discovery approaches' limitations relying on untargeted screening methods, there is a growing trend to employ unconventional secondary metabolomics techniques. Aided by the more in-depth understanding of different biosynthetic pathways and the technological advancement in analytical instrumentation, the development of new methodologies provides an alternative that can accelerate discoveries of new lead-structures of natural origin. This present mini-review briefly discusses selected examples regarding advancements in bioinformatics and genomics (focusing on genome mining and metagenomics approaches), as well as bioanalytics (mass-spectrometry) towards the microbial NPs-based drug discovery and development. The selected recent discoveries from 2015 to 2020 are featured herein.

Keywords: drug discovery; natural products; bioinformatics; genome mining; bioanalytics

1. Introduction

Natural products (NPs) originating from plants, animals, marine organisms, and particularly from microbial sources continue to inspire novel discoveries in chemistry [1], biology [2], and medicine [3]. They possess immense structural and chemical diversity with a wide variety of biological properties. Most pharmacologically relevant antimicrobial, antiviral, anti-inflammatory and analgesic, and antitumor agents and approved small molecule drugs have either been NPs, their derivatives, synthetic compounds with NP pharmacophore, or their synthetic mimics. Notably, more than half of the new small molecule drugs have been developed from microbial NPs over the past decades [4,5].

Current interest in the discovery of NPs, especially from microbial sources, is mostly due to the failure of synthetic libraries to generate the expected number of developmental drug candidates in the pharmaceutical industry during the past 20–30 years. Additionally, the emergence of clinically relevant pathogens becoming increasingly resistant to currently used anti-infectives, i.e., antibiotics, warrants the search for novel bioactive metabolites in the field of microbial NPs [5–9]. However, finding novel NPs has become more difficult as the rediscovery of known NPs is still an increasing challenge. A high rate of the discovery of NPs was yielded by classical methods that recover only a fraction or even none of the desired secondary metabolites. The sharp decline in discoveries arose with limitations of the traditional top-down screening approaches. Those approaches, including bioassay-and chemical signature-guided isolations, have largely been exhausted and may no longer be capable of delivering novel lead compounds [10].

In the search for alternative methods, advancements made in bioinformatics and chemical analysis might hold the key to lead a renaissance in the field of microbial NP discovery. The growing knowledge of different biosynthetic machinery, drug targets, and resistance



Citation: Sukmarini, L. Recent Advances in Discovery of Lead Structures from Microbial Natural Products: Genomics- and Metabolomics-Guided Acceleration. *Molecules* 2021, 26, 2542. https://doi.org/10.3390/ molecules26092542

Academic Editors: Shugeng Cao, Mostafa Rateb and Jianguang Luo

Received: 15 February 2021 Accepted: 22 April 2021 Published: 27 April 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/). mechanisms has served as a launch platform to a new era in the methodological approach for drug discovery [11,12]. Given the rising limitations imposed by uncultivable strains and silent gene clusters, the integrative approach of bottom-up targeted screening, employing advanced analytical methods and guided by bioinformatics analysis, provides a promising alternative for unlocking the microbial metabolomes on an unprecedented scale. This approach eventually leads to disclosing the potential of microbial NP discovery [13–15]. This mini-review highlights in particular some of the most recent advances in microbial NP discoveries as well as their discovery examples in the last five years achieved by the use of genomic and metabolomic approaches. In terms of this, a genomic strategy uncovers the large number of microbial clustered genes (biosynthetic gene clusters) that encode the proteins responsible for the biosynthesis of a new NP that is undetected under standard fermentation conditions, while a metabolomics method embraces the global measurement of small-molecule metabolites from a microbe.

2. Bioinformatics- and Genomics-Driven Discovery

Genomics and metagenomics (which has also been described as environmental genomics, relating to the genomic DNA from an environmental sample) revealed the remarkable biosynthetic potential of microbial NPs and their vast chemical inventory that can be prioritized and systematically mined for novel or new secondary metabolites with desirable bioactivities. The growing application of bioinformatics into a standard practice in discovery projects has varied approaches to identify novel lead structures [16]. Herein, advances in genomics-driven NPs discovery covering bioinformatics-guided identification of biosynthetic gene clusters (BGCs) in (meta)genomes are briefly highlighted. Additionally, the application of innovative technology in situ cultivation in novel compound discovery is also included.

2.1. Genome Mining Approach

Fueled by the fast development of genome sequencing technologies, genome mining evolved during the last decades and is currently an essential part of drug discovery efforts. The genome mining approach detects and analyzes the BGCs of the chemical compounds automatically (computationally) and subsequently connects those genes to molecules. Furthermore, the significance of this approach associated with other techniques leading to drug discovery, especially of microbial NP origin, has been extensively described elsewhere [17– 20]. Although the genome mining approach showcases the full biosynthetic potential of a strain, it is not very worthwhile without linking the predicted secondary metabolite BGCs to their product. Moreover, to take full advantage of NP diversity, BGCs must be prioritized by product novelty or function. BGCs hold the key information to understanding and predicting a specific or a group of related metabolites. By identifying open reading frames (ORFs) in a gene sequence, one can set the borders of the protein-encoding genes, and therein the protein sequence can be predicted through bioinformatics tools. As in some cases, bioinformatics can reveal BGCs with high similarities as a fast evaluation for target novelty; consequently, the time invested with computational work would save extensive resources and efforts only to re-isolate a previously described compound [15,20–23].

As mentioned above, progressions made in bioinformatics are mainly owed to advancements in genomics. Hence, the wealth of genomic information has led to the development of multiple bioinformatics-guided genome mining tools that examine this genomic data to detect and annotate potential BGCs automatically. Nevertheless, the realization of the full potentials of bioinformatics is bound to improvements in information algorithms towards knowledge of NP biosynthetic machinery (e.g., ribosomally synthesized and post-translationally modified peptides/RiPPs and non-ribosomal peptide synthase/NRPS, and polyketide synthase/PKS) [24–27]. Several widely used online platforms are still in active development, as listed in Table 1. Many of these selected tools have been extensively reviewed [21,28–34].

Platform	Description	Web Server URL	Reference
BIG-FAM	Global biosynthetic space of microbial BGC families database	https://bigfam. bioinformatics.nl, accessed on 10 February 2021	Kautsar et al. [35]
MIBiG 2.0	Minimum information on biosynthetic gene clusters (MIBiG) standard respiratory of characterized BGCs	https://mibig. secondarymetabolites. org, accessed on 10 February 2021	Kautsar et al. [36]
antiSMASH 5.0	Automated pipeline to mine secondary metabolite BGCs	https://antismash. secondarymetabolites. org, accessed on 10 February 2021	Blin K et al. [37]
PRISM 4	Automated pipeline to mine secondary metabolite BGCs	http://prism. adapsyn.com, accessed on 10 February 2021	Skinnider et al. [38]
BAGEL 4	Mining of RiPP and bacteriocins BGCs	http://bagel4 .molgenrug.nl, accessed on 10 February 2021	Van Heel et al. [39]
BiG-SPACE - CORASON	Biosynthetic gene similarity clustering and prospecting engine	https: //bigscape-corason. secondarymetabolites. org, accessed on 10 February 2021	Navarro-Muňoz et al. [40]
ARTS	Mining of BGCs on the basis of the prediction of antimicrobial resistance genes that are part of BGCs	https://arts. ziemertlab.com, accessed on 10 February 2021	Alanjary et al. [41]
CASSIS/SMIPS	Mine for PKS, NRPS, and DMATS anchor genes (SMIPS) in fungal genomes; predict gene clusters around anchor genes on the basis of conserved promoter regions	https://sbi.hki-jena. de/cassis/, accessed on 10 February 2021	Wolf et al. [42]
IMG-ABC	A comprehensive database of secondary metabolite BGCs	https://img.jgi.doe. gov/abc, accessed on 10 February 2021	Hadjithomas et al. [43]
RiPPMinner	Analysis of RiPP precursor peptides to predict structural features	https://www.nii.ac. in/~priyesh/ lantipepDB/ newpredictions/ index.php, accessed on 10 February 2021	Agrawal et al. [44]
RiPP-RODEO	Mining and analysis of RiPPs	https://www. ripprodeo.org/, accessed on 10 February 2021	Tiez et al. [45]

Table 1. Selected latest bioinformatics tools dedicated to genome mining NPs (2015–2020).

Streptocollin [46], stackepeptins [47], bicereucin [48], curacozole [49], and lexapeptide [50] (Figure 1) are the recent examples of RiPPs in which their BGCs were discovered using the genome mining approach. These metabolites were successfully characterized by heterologous expression and monitoring production in tandem mass spectrometry (MS) experiments.



Figure 1. Selected recent examples of RiPPs and their BGCs discovering by the genome mining approach. The structural genes within BGC-encoded precursor peptides are depicted in black.

Moreover, two new NRPS-PKS hybrids, guangnanmycin A and weishanmycin A1, were discovered through BGC genome mining of promising anticancer drug leads leinamycin NP family (Figure 2A) [51], while five out of six new NRPS-PKS polycyclic tetramate macrolactams (Figure 2B) from the genome of *Streptomyces* sp. SCSIO 40,010 were identified to have cytotoxic activity [52]. In addition, a broad-spectrum antibacterial of rare sulfur-containing phosphate argolaphos B (Figure 2C) was discovered by mining the genomes of 10,000 actinomycetes [53]. Another example is the thiotetronic acid antibiotics of a new thiolactamycin analog (11-methyl-thiolactomycin) and thiotetroamides A and B (Figure 2D) discovered by a resistance-directed genome mining strategy [54]. By targeting BGCs with duplicated housekeeping genes that may encode protein targets, one is now able to infer the target of uncharacterized NP by analyzing BGC-associated self-resistance genes without prior knowledge of the structure.



Figure 2. Selected recent examples of NRPS-PKS hybrids (**A**,**D**), polycyclic tetramate macrolactam (**B**), and phosphonate family of NPs (**C**) and their biosynthetic gene cluster discovery using the genome mining approach.

2.2. Culture-Independent Strategies and Revolution in Metagenomics

It has been estimated that less than 1% of the bacteria present in most environmental samples are readily susceptible to cultivation using current fermentation technologies. Moreover, 5% of fungal species have been described, while many remain understudied despite their significant source of bioactive metabolites. Extensive studies of microbial 16S rRNA have revealed that the natural diversity of the prokaryotes by far exceed the

number of bacteria that have been described to date. Therefore, in an attempt to decipher novel bioactive metabolites from unidentified microbes, researchers have explored several culture-independent approaches, including the current diffusion chamber technology, isolation chip (iChip). This multichannel device allows for the diffusion of nutrients and growth factors through the chambers. It enables the growth of uncultured bacteria in their natural environment. The application of this technology has led to the discovery of a novel depsipeptide antibiotic teixobactin from a previously unculturable β -proteobacteria named Eleftheria terrae. Interestingly, this antibiotic has displayed no detectable resistance thus far. The BGC identification using a homology search revealed that teixobactin consists of two large NRPS-encoding genes [55–57].

Moreover, the other culture-independent approach of metagenomics has also been established. Metagenomics relies on sampling environmental DNA (eDNA) and assessing their metabolomics independent from the producing organism. This has great implications when considering strains challenging to isolate or cultivate, such as strains from extreme environments and symbionts of marine organisms [16,58]. The revolution in this approach encompassing the phenotypic and homology DNA screening strategies in situ has been ameliorated by the advancement of next-generation sequencing (NGS) technologies [59–61]. By directly capturing eDNA from the environment and subsequently identifying, isolating, and expressing BGCs in a heterologous host, metagenomics has the potential to bring biosynthetic diversity from the environment into drug discovery pipelines.

A study by Brady and his co-workers [62] employing targeted metagenomics of soil samples from different geographic regions led to the discovery of two new antifungal structures belonging to the rare class of tryptophan dimers NPs, hydroxysporine and reductasporine (Figure 3A). Soil samples were pre-screened to identify the most phylogenetically unique CPAS (responsible for the dimerization of activated Trp) gene sequences. Molecules associated with this gene were accessed through targeted metagenomic library construction and heterologous expression in *S. albus* or *E. coli*. Moreover, a class of calcium-dependent antibiotics called malacidins (Figure 3B) was recently discovered by the metagenomics approach of 2000 unique soils. These antibiotics exhibited activity against multidrug-resistant pathogens and sterilized methicillin-resistant *Staphylococcus aureus* [63]. The cyclic lipopeptides of malacidins A and B contain eight amino acids macrocycles and polyunsaturated lipid, incorporating a rare 3-hydroxyl aspartic acid. Another recent breakthrough discovery was the finding of antiviral peptide divamide A (Figure 3C) exhibiting activity against the human immune virus infection. These compounds were synthesized by symbiotic cyanobacteria *Prochloron didemni* living in marine tunicate *Didemnum molle* E11-036 [64].

A.

The hys and red gene clusters (expressed in S. albus and E. coli, respectively)





B. The malacidin BGC (expressed in S. albus)





Figure 3. Selected recent examples of NP discovery employing the metagenomic approach. (**A**) The *hys* and *red* gene clusters containing Trp dimerization enzyme sequence tag (depicted in green), which were expressed in *S. albus* and *E. coli*, respectively, afforded two new rare tryptophan dimers, hydroxysporine and reductasporine. (**B**) The malacidin BGC containing the Asp4 (the domain responsible for incorporating the first aspartic acid) gene sequence that was expressed in *S. albus* led to the production of malacidins A and B containing a rare 3-hydroxyl aspartic acid moiety (HyAsp, highlighted in purple). (**C**) The metagenome sequencing of the DNA sample of whole-tunicate *D. molle* E11-036 revealed the BGC of *divamide* A (the core peptide divamide A is shown in black).

3. Technological Advancements in Bioanalytics: Mass Spectrometry-Based Metabolomics

The key step in compound detection and identification relies directly on analytical instrumentation and data processing software for increased sensitivity and accuracy. Given the need for increased sensitivity in metabolomics, mass spectrometry (MS) is a predominant analytical technique with wide applicability in high-throughput screening programs. It has the potential to uncover elemental composition; structural information, i.e., mass-to-charge ratios (m/z); isotopic patterns; and abundance, as well as fragmentation patterns of molecules. Current separation techniques, including high-performance liquid chromatography (HPLC) or ultra-high-pressure liquid chromatography (UPLC), as well as

gas chromatography (GC), are routinely coupled to MS towards efficient detectability of the generated ions. This coupled system has proved a powerful technique that has contributed towards metabolic profiling [65–67]. It has been known that chemical and electron impact ionization (EI/CI) frequently used with GC–MS and the more recent electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) allow for the analysis of complex molecules such as proteins and peptides. Moreover, the mass analyzer has been developed to employ various detectors, including the time-of-flight analyzer (TOF), the quadrupole ion trap (QIT), the ion cyclotron resonance (ICR), and the orbitrap. While the single-stage MS technique mainly reveals the mass compound, the fragmentation through collision-induced dissociation (CID) for tandem MS (MS/MS or MSⁿ) and during electron ionization (EI) provides the building blocks used to characterize molecules and study their fragmentation behavior. The interpretation and in-depth analysis of these molecular fragments towards accurate identification of NP compounds have been made possible by recent MS technique advances [68–71].

Recent advances in MS that integrate molecular networking (MN) of the MS/MS data have allowed for more rapid dereplication of known molecules from complex mixtures (Figure 4), which in turn have enabled not only the identification of related analogs but also contributed towards unraveling novel compounds by avoiding re-isolation of known compounds. It can be used to explore thousands to millions (and potentially billions) of MS/MS spectra without any prior knowledge regarding the chemical composition of samples. An open-access MN platform Global Natural Products Social Molecular Networking (GNPS; http://gnps.ucsd.edu, accessed on 10 February 2021), can automatically perform a spectral library search for known molecules (if available in public MS/MS spectral libraries) [72,73]. Furthermore, Allard et al. (2016) [74] integrated MN and an extensive in silico MS/MS database, offering a more powerful tool to navigate through the chemistry of complex NP extracts, dereplicate metabolites, and annotate analogs.



Figure 4. Schematic of molecular networking (MN)-based dereplication.

Application of the MN approach to marine microbial *Vibrio* strains has led to the discovery of a series of antibacterial polyketide vitroprocines A-J (Figure 5A) [75] and anti-inflammatory and analgesic sphongonucleosides (Figure 5B) [76]. Recently, MN has been coupled with genome mining to dig more into the BGCs responsible for metabolite production. This method may also be applied in elucidating biosynthetic pathways and conjugation with stable-isotope labeling by amino acids in cell culture (SILAC) in order to provide more comprehensive insights into metabolomics studies, e.g., NPRS-PKS nidulin A [77] and colibactin [78,79]. The information provided by the MN–BGC correlation apparently can be exploited to augment discovery, isolation, and structural prediction of novel compounds produced by an organism, including a microbial strain [80,81]. An association between genomics and metabolomics data allowed for the detection of three new antibiotic NPs, columbamides A, B, and C [82], and a new type of thiomarinol [83] from marine bac-

teria. Additionally, MS-guided genome mining called metabologenomics detects new NPs and connects them with their BGCs. Matched BGC sequence information may be harnessed to elucidate compound structures further and/or to identify additional molecular features for searching. Metabolomics works by grouping similar BGCs from diverse bacteria into gene cluster families (GCFs) [16,84-86]. It should be noted that peptide-based NP discovery has primarily employed this method (peptidogenomics) due to its well-characterized biosynthetic machinery. Non-ribosomal peptide (NRP) tambromycin [87] and the hybrid NRPS-PKS rimosamides [88] are examples of novel NPs detected by the metabologenomic approach (Figure 5C). Through metabologenomic workflow of a 178-strain actinomycetes dataset applying scoring metrics to identify correlations between NP and GCF, these peptides were successfully afforded. Furthermore, a recent discovery of NRP tyrobetaines (Figure 5C) utilizing this workflow in combination with MN showed the great potential of MN-based metabologenomics for identifying novel NPs [89]. The approach has also been extended to the discovery of glycosylated NPs (glycogenomics) such as the marine-derived antibiotic rosamicin derivative and salinipyrone A and pacificanone A (Figure 5D) [90]. By matching tandem MS spectra of a marine bacterium Salinispora pacifica SNS237 with the BGC of type I PKS encoding desosamine (deoxysugar) biosynthesis, research revealed several rosamicin derivatives. Interestingly, mutagenesis experiments have revealed that salinipyrone and pacificanone seem to be by-products of the rosamicin PKS. Moreover, both peptidogenomic and glycogenomic approaches, as well as metabologenomics, have been extensively reviewed very recently elsewhere [81].



Figure 5. Selected recent examples of NPs discovered by the MN-based approach and metabologenomics. (A) Polyketides. (B) Sphongonucleosides. (C) NRP and the hybrid NRPS-PKS. (D) Glycosylated NPs.

4. Conclusions

Remarkably, advancements in bioinformatics tools, genomics, and bioanalytics (particularly in MS) have recently enhanced the field of microbial NP research. These strategies outlined above offer alternatives to accelerate NP drug discovery over conventional methods efficiently. With continued significant progress in both genomics and metabolomics approaches and/or combined with synthetic biology, the microbial NPs discovery field shows strong signs of developing and is ready to lead at the forefront of delivering drugs or drug leads.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The author would like to thank the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) for their support through the Returning Experts-Migration and Diaspora Programme.

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

References

- 1. Trost, B.M.; Wang, Y.; Buckl, A.K.; Huang, Z.; Nguyen, M.H.; Kuzmina, O. Total Synthesis of Bryostatin 3. *Science* 2020, *368*, 1007–1011. [CrossRef] [PubMed]
- Yasuda-Yamahara, M.; Kume, S.; Maegawa, H. Roles of mTOR in Diabetic Kidney Disease. Antioxidants 2021, 10, 321. [CrossRef] [PubMed]
- Dugger, S.A.; Platt, A.; Goldstein, D.B. Drug Development in the Era of Precision Medicine. *Nat. Rev. Drug Discov.* 2018, 17, 183–196. [CrossRef] [PubMed]
- 4. 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]
- Demain, A.L. Importance of Microbial Natural Products and the Need to Revitalize their Discovery. J. Ind. Microbiol. Biotechnol. 2014, 41, 185–201. [CrossRef] [PubMed]
- 6. Kirst, A.H. Developing New Antibacterials through Natural Product Research. *Expert Opin. Drug Discov.* **2013**, *8*, 479–493. [CrossRef] [PubMed]
- Katz, L.; Baltz, R.H. Natural Product Discovery: Past, Present, and Future. J. Ind. Microbiol. Biotechnol. 2016, 43, 155–176. [CrossRef]
- 8. Shen, B. A New Golden Age of Natural Products Drug Discovery. *Cell* **2015**, *163*, 1297–1300. [CrossRef]
- 9. Mathur, S.; Hoskins, C. Drug Development: Lessons from Nature. Biomed. Rep. 2017, 6, 612–614. [CrossRef]
- Pye, C.R.; Bertin, M.J.; Lokey, R.S.; Gerwick, W.H.; Linington, R.G. Retrospective Analysis of Natural Products Provides Insights for Future Discovery Trends. *Proc. Natl. Acad. Sci.USA* 2017, 114, 5601–5606. [CrossRef]
- 11. Johnston, C.W.; Skinnider, A.M.; DeJong, A.C.; Rees, P.N.; Chen, G.M.; Walker, C.G.; French, S.; Brown, E.D.; Bérdy, J.; Liu, D.Y.; et al. Assembly and Clustering of Natural Antibiotics Guides Target Identification. *Nat. Chem. Biol.* 2016, *12*, 233–239. [CrossRef]
- 12. Kohanski, M.A.; Dwyer, D.J.; Collins, J.J. How Antibiotics Kill Bacteria: From Targets to Networks. *Nat. Rev. Genet.* 2010, *8*, 423–435. [CrossRef] [PubMed]
- 13. Harvey, A.L.; Edrada-Ebel, R.; Quinn, R.J. The Re-Emergence of Natural Products for Drug Discovery in the Genomics Era. *Nat. Rev. Drug Discov.* 2015, *14*, 111–129. [CrossRef] [PubMed]
- 14. Wright, G.D. Unlocking the Potential of Natural Products in Drug Discovery. Microb. Biotechnol. 2019, 12, 55–57. [CrossRef]
- 15. Luo, Y.; Cobb, E.R.; Zhao, H. Recent Advances in Natural Product Discovery. *Curr. Opin. Biotechnol.* **2014**, *30*, 230–237. [CrossRef]
- 16. Zhang, M.M.; Qiao, Y.; Ang, E.L.; Zhao, H. Using Natural Products for Drug Discovery: The Impact of the Genomics Era. *Expert Opin. Drug Discov.* **2017**, *12*, 475–487. [CrossRef] [PubMed]
- Yang, Z.; He, J.; Wei, X.; Ju, J.; Ma, J. Exploration and Genome Mining of Natural Products from Marine Streptomyces. *Appl. Microbiol. Biotechnol.* 2019, 104, 67–76. [CrossRef]
- Albarano, L.; Esposito, R.; Ruocco, N.; Costantini, M. Genome Mining as New Challenge in Natural Products Discovery. *Mar. Drugs* 2020, 18, 199. [CrossRef]
- 19. Sekurova, O.N.; Schneider, O.; Zotchev, S.B. Novel Bioactive Natural Products from Bacteria via Bioprospecting, Genome Mining and Metabolic Engineering. *Microb. Biotechnol.* **2019**, *12*, 828–844. [CrossRef]
- Ziemert, N.; Weber, T.; Medema, M.H. Genome Mining Approaches to Bacterial Natural Product Discovery. In *Comprehensive Natural Products III*; Elsevier BV: Amsterdam, The Netherlands, 2019; Volume 6, pp. 19–33.
- Ziemert, N.; Alanjary, M.; Weber, T. The Evolution of Genome Mining in Microbes—A Review. Nat. Prod. Rep. 2016, 33, 988–1005. [CrossRef]
- Kalkreuter, E.; Pan, G.; Cepeda, A.J.; Shen, B. Targeting Bacterial Genomes for Natural Product Discovery. *Trends Pharmacol. Sci.* 2020, 41, 13–26. [CrossRef] [PubMed]
- 23. Wohlleben, W.; Mast, Y.; Stegmann, E.; Ziemert, N. Antibiotic Drug Discovery. Microb. Biotechnol. 2016, 9, 541–548. [CrossRef]

- Wang, H.; Fewer, D.P.; Holm, L.; Rouhiainen, L.; Sivonen, K. Atlas of Nonribosomal Peptide and Polyketide Biosynthetic Pathways Reveals Common Occurrence of Nonmodular Enzymes. *Proc. Natl. Acad. Sci. USA* 2014, 111, 9259–9264. [CrossRef] [PubMed]
- Cimermancic, P.; Medema, M.H.; Claesen, J.; Kurita, K.; Brown, L.C.W.; Mavrommatis, K.; Pati, A.; Godfrey, P.A.; Koehrsen, M.; Clardy, J.; et al. Insights into Secondary Metabolism from a Global Analysis of Prokaryotic Biosynthetic Gene Clusters. *Cell* 2014, 158, 412–421. [CrossRef] [PubMed]
- 26. Scott, T.A.; Piel, J. The Hidden Enzymology of Bacterial Natural Product Biosynthesis. Nat. Rev. Chem. 2019, 3, 404–425. [CrossRef]
- 27. Moghaddam, J.A.; Jautzus, T.; Alanjary, M.; Beemelmanns, C. Recent Highlights of Biosynthetic Studies on Marine Natural Products. *Org. Biomol. Chem.* **2021**, *19*, 123–140. [CrossRef]
- 28. Weber, T.; Kim, H.U. The Secondary Metabolite Bioinformatics Portal: Computational Tools to Facilitate Synthetic Biology of Secondary Metabolite Production. *Synth. Syst. Biotechnol.* **2016**, *1*, 69–79. [CrossRef]
- Lee, N.; Hwang, S.; Kim, J.; Cho, S.; Palsson, B.; Cho, B.-K. Mini Review: Genome Mining Approaches for the Identification of Secondary Metabolite Biosynthetic Gene Clusters in Streptomyces. *Comput. Struct. Biotechnol. J.* 2020, 18, 1548–1556. [CrossRef]
- Nett, M. Genome Mining: Concept and Strategies for Natural Product Discovery. In *Progress in the Chemistry of Organic Natural Products;* Kinghorn, A.D., Falk, H., Kobayashi, J., Eds.; Springer International Publishing: Cham, Switzerland, 2014; Volume 99, pp. 199–245.
- Medema, M.H.; Fischbach, M.A. Computational Approaches to Natural Product Discovery. Nat. Chem. Biol. 2015, 11, 639–648. [CrossRef]
- Kim, H.U.; Blin, K.; Lee, S.Y.; Weber, T. Recent Development of Computational Resources for New Antibiotics Discovery. Curr. Opin. Microbiol. 2017, 39, 113–120. [CrossRef]
- Russell, A.H.; Truman, A.W. Genome Mining Strategies for Ribosomally Synthesised and Post-Translationally Modified Peptides. Comput. Struct. Biotechnol. J. 2020, 18, 1838–1851. [CrossRef] [PubMed]
- 34. Kloosterman, A.M.; Medema, M.H.; Van Wezel, G.P. Omics-Based Strategies to Discover Novel Classes of RiPP Natural Products. *Curr. Opin. Biotechnol.* 2021, 69, 60–67. [CrossRef]
- 35. Kautsar, S.A.; Blin, K.; Shaw, S.; Weber, T.; Medema, M.H. Big-Fam: The Biosynthetic Gene Cluster Families Database. *Nucleic Acids Res.* 2020, 49, D490–D497. [CrossRef]
- Kautsar, A.S.; Blin, K.; Shaw, S.; Navarro-Muñoz, J.C.; Terlouw, B.R.; Van Der Hooft, J.J.J.; Van Santen, A.J.; Tracanna, V.; Duran, H.G.S.; Andreu, V.P.; et al. Mibig 2.0: A Repository for Biosynthetic Gene Clusters of Known Function. *Nucleic Acids Res.* 2019, 48, D454–D458. [CrossRef]
- 37. 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. [CrossRef]
- Skinnider, M.A.; Johnston, C.W.; Gunabalasingam, M.; Merwin, N.J.; Kieliszek, A.M.; MacLellan, R.J.; Li, H.; Ranieri, M.R.M.; Webstar, A.L.H.; Cao, M.P.T.; et al. Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. *Nat. Commun.* 2020, 6058. [CrossRef]
- 39. Van Heel, A.J.; De Jong, A.; Song, C.; Viel, J.H.; Kok, J.; Kuipers, O.P. BAGEL4: A User-Friendly Web Server to thoroughly Mine RiPPs and Bacteriocins. *Nucleic Acids Res.* 2018, 46, W278–W281. [CrossRef]
- Navarro-Muñoz, J.C.; Selem-Mojica, N.; Mullowney, M.W.; Kautsar, S.A.; Tryon, J.H.; Parkinson, E.I.; De Los Santos, E.L.; 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]
- Alanjary, M.; Kronmiller, B.; Adamek, M.; Blin, K.; Weber, T.; Huson, D.H.; Philmus, B.; Ziemert, N. The Antibiotic Resistant Target Seeker (ARTS), an Exploration Engine for Antibiotic Cluster Prioritization and Novel Drug Target Discovery. *Nucleic Acids Res.* 2017, 45, W42–W48. [CrossRef] [PubMed]
- 42. Wolf, T.; Shelest, V.; Nath, N.; Shelest, E. CASSIS and SMIPS: Promoter-Based Prediction of Secondary Metabolite Gene Clusters in Eukaryotic Genomes. *Bioinformatics* **2015**, *32*, 1138–1143. [CrossRef] [PubMed]
- Hadjithomas, M.; Chen, I.-M.A.; Chu, K.; Huang, J.; Ratner, A.; Palaniappan, K.; Andersen, E.; Markowitz, V.; Kyrpides, N.C.; Ivanova, N.N. IMG-ABC: New Features for Bacterial Secondary Metabolism Analysis and Targeted Biosynthetic Gene Cluster Discovery in Thousands of Microbial Genomes. *Nucleic Acids Res.* 2017, 45, D560–D565. [CrossRef]
- 44. Agrawal, P.; Khater, S.; Gupta, M.; Sain, N.; Mohanty, D. RiPPMiner: A Bioinformatics Resource for Deciphering Chemical Structures of RiPPs based on Prediction of Cleavage and Cross-Links. *Nucleic Acids Res.* **2017**, *45*, W80–W88. [CrossRef]
- 45. Tietz, J.I.; Schwalen, C.J.; Patel, P.S.; Maxson, T.; Blair, P.M.; Tai, H.-C.; Zakai, U.I.; Mitchell, D.A. A New Genome-Mining Tool Redefines the Lasso Peptide Biosynthetic Landscape. *Nat. Chem. Biol.* **2017**, *13*, 470–478. [CrossRef]
- Iftime, D.; Jasyk, M.; Kulik, A.; Imhoff, J.F.; Stegmann, E.; Wohlleben, W.; Süssmuth, R.D.; Weber, T. Streptocollin, a Type IV Lanthipeptide Produced byStreptomyces collinusTü 365. *ChemBioChem* 2015, *16*, 2615–2623. [CrossRef] [PubMed]
- 47. Jungmann, N.A.; Van Herwerden, E.F.; Hügelland, M.; Süssmuth, R.D. The Supersized Class III Lanthipeptide Stackepeptin Displays Motif Multiplication in the Core Peptide. *ACS Chem. Biol.* **2015**, *11*, 69–76. [CrossRef] [PubMed]
- 48. Huo, L.; van der Donk, W.A. Discovery and Characterization of Bicereucin, an Unusual d-Amino Acid-Containing Mixed Two-Component Lantibiotic. *J. Am. Chem. Soc.* 2016, 138, 5254–5257. [CrossRef]

- Kaweewan, I.; Komaki, H.; Hemmi, H.; Hoshino, K.; Hosaka, T.; Isokawa, G.; Oyoshi, T.; Kodani, S. Isolation and Structure Determination of a New Cytotoxic Peptide, Curacozole, from Streptomyces Curacoi based on Genome Mining. *J. Antibiot.* 2018, 72, 1–7. [CrossRef] [PubMed]
- Xu, M.; Zhang, F.; Cheng, Z.; Bashiri, G.; Wang, J.; Hong, J.; Wang, Y.; Xu, L.; Chen, X.; Huang, S.; et al. Functional Genome Mining Reveals a Class V Lanthipeptide Containing a d -Amino Acid Introduced by an F 420 H 2 -Dependent Reductase. *Angew. Chem.* 2020, 132, 18029–18035. [CrossRef]
- 51. Pan, G.; Xu, Z.; Guo, Z.; Hindra; Ma, M.; Yang, D.; Zhou, H.; Gansemans, Y.; Zhu, X.; Huang, Y.; et al. Discovery of the Leinamycin Family of Natural Products by Mining Actinobacterial Genomes. *Proc. Nat. Acad. Sci. USA* **2017**, *114*, E11131–E11140. [CrossRef]
- Liu, W.; Zhang, W.; Jin, H.; Zhang, Q.; Chen, Y.; Jiang, X.; Zhang, G.; Zhang, L.; Zhang, W.; She, Z.; et al. Genome Mining of Marine-Derived Streptomyces sp. SCSIO 40010 Leads to Cytotoxic New Polycyclic Tetramate Macrolactams. *Mar. Drugs* 2019, 17, 663. [CrossRef]
- Ju, K.-S.; Gao, J.; Doroghazi, J.R.; Wang, K.-K.A.; Thibodeaux, C.J.; Li, S.; Metzger, E.; Fudala, J.; Su, J.; Zhang, J.K.; et al. Discovery of Phosphonic Acid Natural Products by Mining the Genomes of 10,000 Actinomycetes. *Proc. Natl. Acad. Sci. USA* 2015, 112, 12175–12180. [CrossRef] [PubMed]
- Tang, X.; Li, J.; Millán-Aguiñaga, N.; Zhang, J.J.; O'Neill, E.C.; Ugalde, J.A.; Jensen, P.R.; Mantovani, S.M.; Moore, B.S. Identification of Thiotetronic Acid Antibiotic Biosynthetic Pathways by Target-Directed Genome Mining. ACS Chem. Biol. 2015, 10, 2841–2849. [CrossRef] [PubMed]
- 55. Ling, L.L.; Schneider, T.; Peoples, A.J.; Spoering, A.L.; Engels, I.; Conlon, B.P.; Mueller, A.; Schäberle, T.F.; Hughes, D.E.; Epstein, S.; et al. A New Antibiotic Kills Pathogens without Detectable Resistance. *Nature* **2015**, *517*, 455–459. [CrossRef]
- 56. Berdy, B.; Spoering, A.L.; Ling, L.L.; Epstein, S.S. In Situ Cultivation of Previously Uncultivable Microorganisms using the Ichip. *Nat. Protoc.* **2017**, *12*, 2232–2242. [CrossRef] [PubMed]
- 57. Shukla, R.; Medeiros-Silva, J.; Parmar, A.; Vermeulen, B.J.A.; Das, S.; Paioni, A.L.; Jekhmane, S.; Lorent, J.; Bonvin, A.M.J.J.; Baldus, M.; et al. Mode of Action of Teixobactins in Cellular Membranes. *Nat. Commun.* **2020**, *11*, 1–10. [CrossRef] [PubMed]
- Charlop-Powers, Z.; Milshteyn, A.; Brady, S.F. Metagenomic Small Molecule Discovery Methods. *Curr. Opin. Microbiol.* 2014, 19, 70–75. [CrossRef]
- 59. Loman, N.J.; Pallen, M.J. Twenty Years of Bacterial Genome Sequencing. Nat. Rev. Genet. 2015, 13, 787–794. [CrossRef]
- 60. Alves, L.D.F.; Westmann, C.A.; Lovate, G.L.; De Siqueira, G.M.V.; Borelli, T.C.; Guazzaroni, M.-E. Metagenomic Approaches for Understanding New Concepts in Microbial Science. *Int. J. Genom.* **2018**, *2018*, 1–15. [CrossRef]
- 61. Datta, S.; Rajnish, K.N.; Samuel, M.S.; Pugazlendhi, A.; Selvarajan, E. Metagenomic Applications in Microbial Diversity, Bioremediation, Pollution Monitoring, Enzyme and Drug Discovery. A Review. *Environ. Chem. Lett.* **2020**, *18*, 1229–1241. [CrossRef]
- 62. Chang, F.-Y.; Ternei, M.A.; Calle, P.Y.; Brady, S.F. Targeted Metagenomics: Finding Rare Tryptophan Dimer Natural Products in the Environment. J. Am. Chem. Soc. 2015, 137, 6044–6052. [CrossRef] [PubMed]
- 63. Hover, B.M.; Kim, S.-H.; Katz, M.; Charlop-Powers, Z.; Owen, J.G.; Ternei, M.A.; Maniko, J.; Estrela, A.B.; Molina, H.; Park, S.; et al. Culture-Independent Discovery of the Malacidins as Calcium-Dependent Antibiotics with Activity against Multi-Drug-Resistant Gram-Positive Pathogens. *Nat. Microbiol.* **2018**, *3*, 415–422. [CrossRef] [PubMed]
- 64. Smith, T.E.; Pond, C.D.; Pierce, E.; Harmer, Z.P.; Kwan, J.; Zachariah, M.M.; Harper, M.K.; Wyche, T.P.; Matainaho, T.K.; Bugni, T.S.; et al. Accessing Chemical Diversity from the Uncultivated Symbionts of Small Marine Animals. *Nat. Chem. Biol.* **2018**, *14*, 179–185. [CrossRef] [PubMed]
- 65. Kind, T.; Fiehn, O. Seven Golden Rules for Heuristic Filtering of Molecular Formulas Obtained by Accurate Mass Spectrometry. BMC Bioinform. 2007, 8, 105. [CrossRef] [PubMed]
- 66. Maurer, H.H. What is the Future of (Ultra) High Performance Liquid Chromatography Coupled to Low and High Resolution Mass Spectrometry for Toxicological Drug Screening? *J. Chromatogr. A* 2013, 1292, 19–24. [CrossRef] [PubMed]
- 67. Hernández, F.H.; Sancho, J.V.; Ibáñez, M.; Abad, E.; Portolés, T.; Mattioli, L. Current Use of High-Resolution Mass Spectrometry in the Environmental Sciences. *Anal. Bioanal. Chem.* **2012**, 403, 1251–1264. [CrossRef]
- Neumann, S.; Böcker, S. Computational Mass Spectrometry for Metabolomics: Identification of Metabolites and Small Molecules. *Anal. Bioanal. Chem.* 2010, 398, 2779–2788. [CrossRef]
- 69. Matsuda, F. Technical Challenges in Mass Spectrometry-Based Metabolomics. Mass Spectrom. 2016, 5, S0052. [CrossRef]
- 70. Scheubert, K.; Hufsky, F.; Böcker, S. Computational Mass Spectrometry for Small Molecules. J. Chemin 2013, 5, 12. [CrossRef]
- 71. Zubarev, R.A.; Makarov, A. Orbitrap Mass Spectrometry. Anal. Chem. 2013, 85, 5288–5296. [CrossRef]
- 72. Yang, J.Y.; Sanchez, L.M.; Rath, C.M.; Liu, X.; Boudreau, P.D.; Bruns, N.; Glukhov, E.; Wodtke, A.; De Felicio, R.; Fenner, A.; et al. Molecular Networking as a Dereplication Strategy. *J. Nat. Prod.* **2013**, *76*, 1686–1699. [CrossRef]
- 73. Quinn, R.A.; Nothias, L.-F.; Vining, O.; Meehan, M.; Esquenazi, E.; Dorrestein, P.C. Molecular Networking As a Drug Discovery, Drug Metabolism, and Precision Medicine Strategy. *Trends Pharmacol. Sci.* **2017**, *38*, 143–154. [CrossRef]
- Allard, P.-M.; Péresse, T.; Bisson, J.; Gindro, K.; Marcourt, L.; Pham, V.C.; Roussi, F.; Litaudon, M.; Wolfender, J.-L. Integration of Molecular Networking and In-Silico MS/MS Fragmentation for Natural Products Dereplication. *Anal. Chem.* 2016, 88, 3317–3323. [CrossRef]

- Liaw, C.-C.; Chen, P.-C.; Shih, C.-J.; Tseng, S.-P.; Lai, Y.-M.; Hsu, C.-H.; Dorrestein, P.C.; Yang, Y.-L. Vitroprocines, New Antibiotics against Acinetobacter Baumannii, Discovered from Marine Vibrio sp. QWI-o6 using Mass-Spectrometry-Based Metab-olomics Approach. *Sci. Rep.* 2015, *5*, 12856. [CrossRef] [PubMed]
- 76. Bertin, M.J.; Schwartz, S.L.; Lee, J.; Korobeynikov, A.; Dorrestein, P.C.; Gerwick, L.; Gerwick, W.H. Spongosine Production by aVibrio Harveyistrain Associated with the SpongeTectitethya crypta. *J. Nat. Prod.* **2015**, *78*, 493–499. [CrossRef] [PubMed]
- 77. Klitgaard, A.; Nielsen, J.B.; Frandsen, R.J.N.; Andersen, M.R.; Nielsen, K.F. Combining Stable Isotope Labeling and Molecular Networking for Biosynthetic Pathway Characterization. *Anal. Chem.* **2015**, *87*, 6520–6526. [CrossRef] [PubMed]
- 78. Vizcaino, M.I.; Crawford, J.M. The Colibactin Warhead Crosslinks DNA. Nat. Chem. 2015, 7, 411–417. [CrossRef]
- 79. Xue, M.; Kim, C.S.; Healy, A.R.; Wernke, K.M.; Wang, Z.; Frischling, M.C.; Shine, E.E.; Wang, W.; Herzon, S.B.; Crawford, J.M. Structure Elucidation of Colibactin and its DNA Cross-Links. *Science* **2019**, *365*, eaax2685. [CrossRef] [PubMed]
- Fox Ramos, A.E.; Evanno, L.; Poupon, E.; Champy, P.; Beniddir, M.A. Natural Products Targeting Strategies Involving Molec-ular Networking: Different Manners, One Goal. *Nat. Prod. Rep.* 2019, *36*, 960–980. [CrossRef]
- Crüsemann, M. Coupling Mass Spectral and Genomic Information to Improve Bacterial Natural Product Discovery Workflows. Mar. Drugs 2021, 19, 142. [CrossRef]
- Kleigrewe, K.; Almaliti, J.; Tian, I.Y.; Kinnel, R.B.; Korobeynikov, A.; Monroe, E.A.; Duggan, B.M.; Di Marzo, V.; Sherman, D.H.; Dorrestein, P.C.; et al. Combining Mass Spectrometric Metabolic Profiling with Genomic Analysis: A Powerful Approach for Discovering Natural Products from Cyanobacteria. J. Nat. Prod. 2015, 78, 1671–1682. [CrossRef]
- Maansson, M.; Vynne, N.G.; Klitgaard, A.; Nybo, J.L.; Melchiorsen, J.; Nguyen, D.D.; Sanchez, L.M.; Ziemert, N.; Dorrestein, P.C.; Andersen, M.R.; et al. An Integrated Metabolomic and Genomic Mining Workflow to Uncover the Biosynthetic Potential of Bacteria. *mSystems* 2016, 1, 00028-15. [CrossRef]
- 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] [PubMed]
- 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] [PubMed]
- 86. Krug, D.; Müller, R. Secondary Metabolomics: The Impact of Mass Spectrometry-Based Approaches on the Discovery and Characterization of Microbial Natural Products. *Nat. Prod. Rep.* **2014**, *31*, 768–783. [CrossRef] [PubMed]
- 87. 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 Nonri-bosomal Peptide with an Unusual Amino Acid Monomer. *ACS Cent. Sci.* **2016**, *2*, 99–108. [CrossRef]
- 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]
- 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] [PubMed]
- Awakawa, T.; Crüsemann, M.; Munguia, J.; Ziemert, N.; Nizet, V.; Fenical, W.; Moore, B.S. Salinipyrone and Pacificanone Are Biosynthetic By-Products of the Rosamicin Polyketide Synthase. *ChemBioChem Eur. J. Chem. Biol.* 2015, 16, 1443–1447. [CrossRef] [PubMed]