

Review

The Deep Mining Era: Genomic, Metabolomic, and Integrative Approaches to Microbial Natural Products from 2018 to 2024

Zhaochao Wang¹, Juanjuan Yu¹, Chenjie Wang¹, Yi Hua^{1,2}, Hong Wang^{1,*}  and Jianwei Chen^{1,*} 

¹ College of Pharmaceutical Science & Collaborative Innovation Center of Yangtze River Delta Region Green Pharmaceuticals, Zhejiang University of Technology, Hangzhou 310014, China; 211122070145@zjut.edu.cn (Z.W.); 221123070219@zjut.edu.cn (J.Y.); 211123070063@zjut.edu.cn (C.W.); huayi@zjut.edu.cn (Y.H.)

² School of Pharmaceutical Sciences, Guangzhou University of Chinese Medicine, Guangzhou 510006, China

* Correspondence: hongw@zjut.edu.cn (H.W.); cjw983617@zjut.edu.cn (J.C.)

Abstract

Over the past decade, microbial natural products research has witnessed a transformative “deep-mining era” driven by key technological advances such as high-throughput sequencing (e.g., PacBio HiFi), ultra-sensitive HRMS (resolution $\geq 100,000$), and multi-omics synergy. These innovations have shifted discovery from serendipitous isolation to data-driven, targeted mining. These innovations have transitioned discovery from serendipitous isolation to data-driven targeted mining. Genome mining pipelines (e.g., antiSMASH 7.0 and DeepBGC) can now systematically discover hidden biosynthetic gene clusters (BGCs), especially in under-explored taxa. Metabolomics has achieved unprecedented accuracy, enabling researchers to target novel compounds in complex extracts. Integrated strategies—combining genomic prediction, metabolomics analysis, and experimental validation—constitute new paradigms of current “deep mining”. This review provides a systematic overview of 185 novel microbial natural products discovered between 2018 and 2024, and dissects how these technological leaps have reshaped the discovery paradigm from traditional isolation to data-driven mining.

Keywords: genomics; metabolomics; secondary metabolites



Academic Editor: Ipek Kurtboke

Received: 29 May 2025

Revised: 20 June 2025

Accepted: 21 June 2025

Published: 23 June 2025

Citation: Wang, Z.; Yu, J.; Wang, C.; Hua, Y.; Wang, H.; Chen, J. The Deep Mining Era: Genomic, Metabolomic, and Integrative Approaches to Microbial Natural Products from 2018 to 2024. *Mar. Drugs* **2025**, *23*, 261. <https://doi.org/10.3390/md23070261>

Copyright: © 2025 by the authors. 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

Natural products (NPs) have long been a key source of therapeutic lead compounds, with microbial natural products alone accounting for approximately 35% of FDA-approved small molecule drugs since 1981 [1,2]. Representative drugs in this category include antibiotics (penicillin), immunomodulators (cyclosporine), metabolic modulators (lovastatin), and others, underscoring their enduring relevance in drug discovery [3–5]. However, the decline of traditional “one-strain, one-compound” screening at the beginning of the 21st century has led to a pronounced bottleneck in microbial natural product discovery, prompting calls for technological innovations to explore untapped chemical diversity [6,7].

Over the past decade, synergistic advances in genomics, metabolomics, and numerous integrative strategies have driven a paradigm shift toward data-driven deep mining [8–11]. PacBio HiFi (long read accuracy > 99.9%) and Nanopore MinION (portable real-time analysis) have enabled microbial comprehensive genome analysis, revealing that only approximately 10% of the biosynthetic gene clusters (BGCs) in *Streptomyces* are expressed under standard culture conditions [12–14]. At the same time, metabolic detection technologies represented by ultrasensitive mass spectrometry and structural analysis technologies

represented by cryogenic nuclear magnetic resonance (600 MHz + magnets), crystal sponge technology, nuclear magnetic resonance (NMR) calculations, and ECD spectroscopy are now capable of detecting trace metabolites and resolving complex stereochemistry, respectively, addressing historical limitations in structural analysis [15–19].

Genome mining enabled the transition from phenotype-driven segregation to computerized prediction, revolutionizing NPs discovery. From the 1970s, with ϕ X174-based Sanger sequencing, to today's evolution into modern bioinformatics techniques utilizing machine learning and comparative genomics, rapid advances in the field of computers have begun to drive the reconfiguration of natural product chemistry technology [20–25]. The basic principle of sequencing is shown in Figure 1. Platforms like anti-smash 7.0 (2023) integrating hidden Markov models (HMMs) and artificial intelligence have expanded the number of annotatable BGC types to more than 40, while DeepBGC (2024) uses bi-directional long and short-term memory networks (BiLSTM) and Random Forests to identify orphan clusters in under-explored phyla (e.g., verrucose microbes) [26–32]. Key to this progress is multi-tool synergies, such as the combination of PRISM 2.0 for ribosomal peptide prediction with ClusterFinder for the analysis of polyketide-non-ribosomal peptide hybrids, which increased structural diversity coverage by 40% compared to single-tool analyses [33]. It is foreseen that next-generation sequencing (NGS) will further accelerate the development of this field [34,35]. If the cost of genome sequencing can be reduced by 99% within 15 years, the realization of large-scale projects like The Global Ocean Microbiome (GOMC) will no longer be limited by the high consumption of funds. Such large-scale metagenomic analyses could very quickly accelerate the discovery of cryptic BGCs encoding novel chemical skeletons, such as cyclopropane-ether hetero-polyketides from soil-derived *Microcystis aeruginosa*, validated via CRISPRi-mediated pathway activation [36].

As opposed to genomics, metabolomics has become another key enabler driven by separation and spectroscopic analysis techniques. The widespread adoption of high-resolution mass spectrometry platforms, including orbital trap, time-of-flight (TOF), and Fourier-transform ion cyclotron resonance (FT-ICR) systems, has dramatically improved detection sensitivity and mass measurement accuracy in metabolomics studies [37–40]. The working principle of various mass spectrometers is shown in Figure 1. Advances have also been made in NMR technology—2D NMR (COSY, HSQC) combined with cryogenic probes has increased signal sensitivity by 30%, enabling the stereochemical identification of marine diterpenes such as eunicellane, a feat unattainable with NMR technology a decade ago [41–44]. The launch of GNPS (Global Natural Products Social Molecular Networking) in 2016 marked a turning point, enabling community-wide sharing of MS/MS data to build metabolite association networks [45]. Combining feature-based molecular networking (FBMN) based on the GNPS platform with artificial intelligence tools (SIRIUS for molecular formula prediction and DeepMass for structure elucidation), researchers can now annotate unknown constituents in *Streptomyces* extracts with up to 65% higher accuracy than database-dependent methods, even for non-model strains [46–48].

The advantage of multi-omics integration in in-depth mining lies in the following: genomics reveals the strain's potential for active product production, metabolomics captures specific secondary metabolites, and the combination of the two realizes a comprehensive analysis of biological systems from genes to phenotypes. Early integration approaches focused on correlation analysis to determine statistical associations between gene expression profiles and metabolite levels [49,50]. Nowadays, after genomics analysis, it is preferred to select high-value BGCs for subsequent product studies, such as the discovery of Mandimycin, through heterologous expression or One Strain Many Compounds (OS-MAC) strategies [51–54]. This integration strategy addresses the “genome-metabolome gap”—where only 25% of predicted BGCs have known products—and strengthens the link

between genomic BGCs data and real chemical objects. In this paper, we categorize the integration strategies into two types: those based on the combination of NMR and genomic analysis, and those based on the combination of liquid chromatograph mass spectrometer (LC-MS) and genomic analysis, show the scope of their applications, and highlight recent advances in combinatorial strategies for mining new NPs.

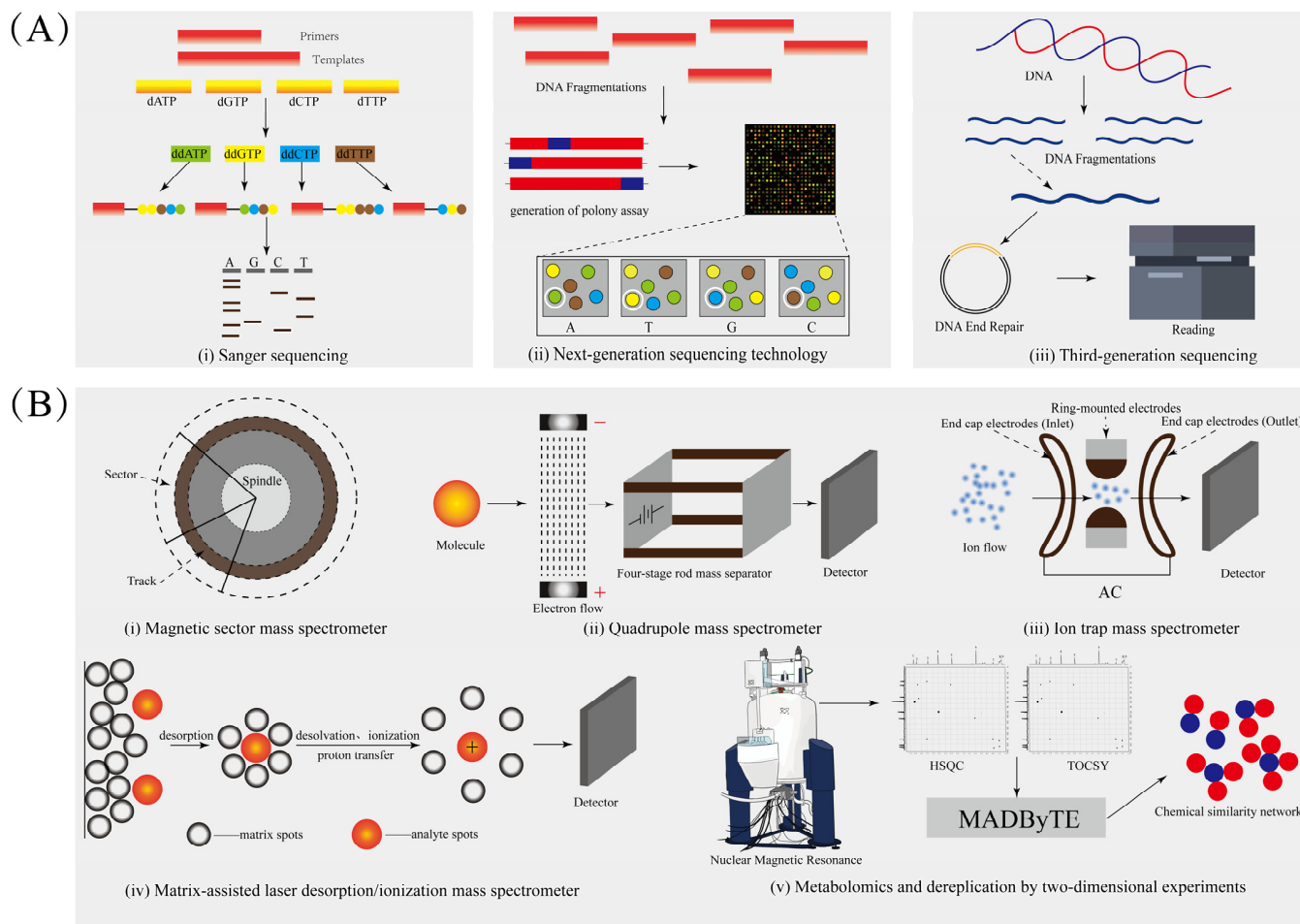


Figure 1. (A) Fundamentals of gene sequencing from the first to the third generation. (i) Sanger sequencing operates on the principle of selective chain termination during DNA replication; (ii) Next-generation sequencing is based on large-scale parallel sequencing and sequencing-by-synthesis; (iii) Third-generation sequencing technologies perform direct single-molecule sequencing of DNA/RNA, featuring long read lengths. (B) The basic principle of action of a variety of mass spectrometers. (i) Magnetic sector mass spectrometer employs magnetic field deflection to differentiate ion masses; (ii) Quadrupole mass spectrometry employs oscillating and static electric fields to selectively transmit ions based on their m/z ratios; (iii) Ion trap technology utilizes oscillating electric fields to spatially confine charged particles, followed by mass-dependent ejection through controlled destabilization of ion trajectories; (iv) Matrix-assisted laser desorption/ionization represents a soft ionization technique in which pulsed laser irradiation of analyte-embedded matrix crystals induces desorption and ionization of intact macromolecules for mass spectrometric detection; (v) The MAD-ByTE platform employs TOCSY and HSQC spectral data to characterize composite spin systems and construct chemical similarity networks across sample cohorts.

2. Genome Mining for NPs

2.1. Genome Mining of RiPPs

Genomic analysis of ribosomally synthesized and post-translationally modified peptides (RiPPs) places significant emphasis on post-translational modifications (PTMs), with RiPP family classification typically determined by their signature PTM enzymes [55]. A

notable development in this field is the emergence of aromatic amino acid-containing RiPPs as a novel subclass, where the formation of specific C-C or C-N bonds during PTMs has been associated with P450 enzymes [56–59]. However, conventional genomics tools lack the capability to establish precise correlations between P450 enzymes and their corresponding RiPP products, necessitating the development of more innovative approaches to systematically identify P450 enzyme-modified RiPPs.

The discovery of P450-associated RiPPs was achieved through an integrated workflow combining three advanced bioinformatics approaches: multilayer sequence similarity network (MSSN) for analyzing functional correlations among biomolecules, short peptide and enzyme co-localization (SPECO) for genome mining of RiPP BGCs, and AlphaFold-Multimer for predicting protein complex structures [60]. This comprehensive strategy began with SPECO analysis of 20,399 actinomycete genomes to identify potential PTM-involved sequences, followed by AlphaFold-Multimer structural predictions that revealed a conserved binding mode where precursor peptides embed their C-termini within P450 pockets while extending core peptides toward the heme center—a molecular signature enabling discrimination between authentic RiPP precursors and non-RiPP sequences. [61]. The workflow further incorporated MSSN construction of precursor peptide–P450 pairs, validated using established tryptorubin A and cittilin A families as reference datasets, which successfully identified three known and three novel RiPP families. Heterologous expression of four selected BGCs (*kst*, *mci*, *scn*, and *sgr*) in *E. coli* yielded five structurally diverse macrocyclic peptides, namely kitasatide 1019 (1), kitasatide 1017 (2), micitide 982 (3), strecintide 839 (4) and gristide 834 (5). The structures of these five compounds as shown in Figure 2. Biochemical characterization revealing that KstB, ScnB, and MciB are substrate-mixed P450 enzymes, which can achieve catalytic cyclization of unnatural precursor peptides and targeted substitution of specific amino acid residues, exhibiting great potential in the field of enzyme engineering. This method utilizes RiPP biosynthetic principles combined with multidimensional sequence-structure analyses to effectively distinguish between RiPPs-associated and non-associated P450s, thereby enhancing the precision and throughput of novel P450 discovery while expanding the target landscape for RiPP research.

The identification of P450-modified RiPPs was facilitated through a streamlined bioinformatics pipeline integrating multiple computational tools. Initial sequence analysis employed BlastP for homologous protein identification and the Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST) for generating sequence similarity networks (SSNs), which enabled BLAST-based similarity analysis and network visualization [62]. RiPPER is a standardized prediction program based on ribosome binding sites (RBSs), specifically designed for predicting, analyzing, and studying RiPPs [63,64]. The rational use of known P450 sequences can also establish the correlation between P450 enzymes and RiPPs. Using three characterized P450 enzymes (BytO, CitB, and TrpB) as query sequences, Hui-Ming Ge and his team performed comprehensive database mining through BlastP searches against NCBI followed by EFI-EST analysis, yielding 13,896 non-redundant P450 sequences after length filtering [65]. The approximate mining process is shown in Figure 3A. RiPPER predicted potential precursor peptide sequences adjacent to P450 enzymes, focusing on those with multiple conserved aromatic amino acids. Combined with SSN analysis, 1057 P450-modified RiPPs gene clusters were successfully identified and classified into 11 categories. Heterologous expression of five novel BGCs, namely *tsu*, *oli*, *san*, *cre*, and *syr*, in *S. albus* J1074 yielded nine new compounds: tsukirubins A–C (6–8), olivorubins A–B (9–10), shandoamide (11), syrinamides A–B (12–13), and citreamide (14), which are shown in Figure 2. This study uncovered 11 new P450-modified RiPPs from four classes via genome mining, broadening structural diversity [65]. A simple and efficient RBS-based workflow

was established for identifying P450-modified RiPPs gene clusters. Unlike complex prediction tools, it utilizes RBS as the identification label and integrates SSN analysis, significantly reducing computational effort while improving accuracy and efficiency.

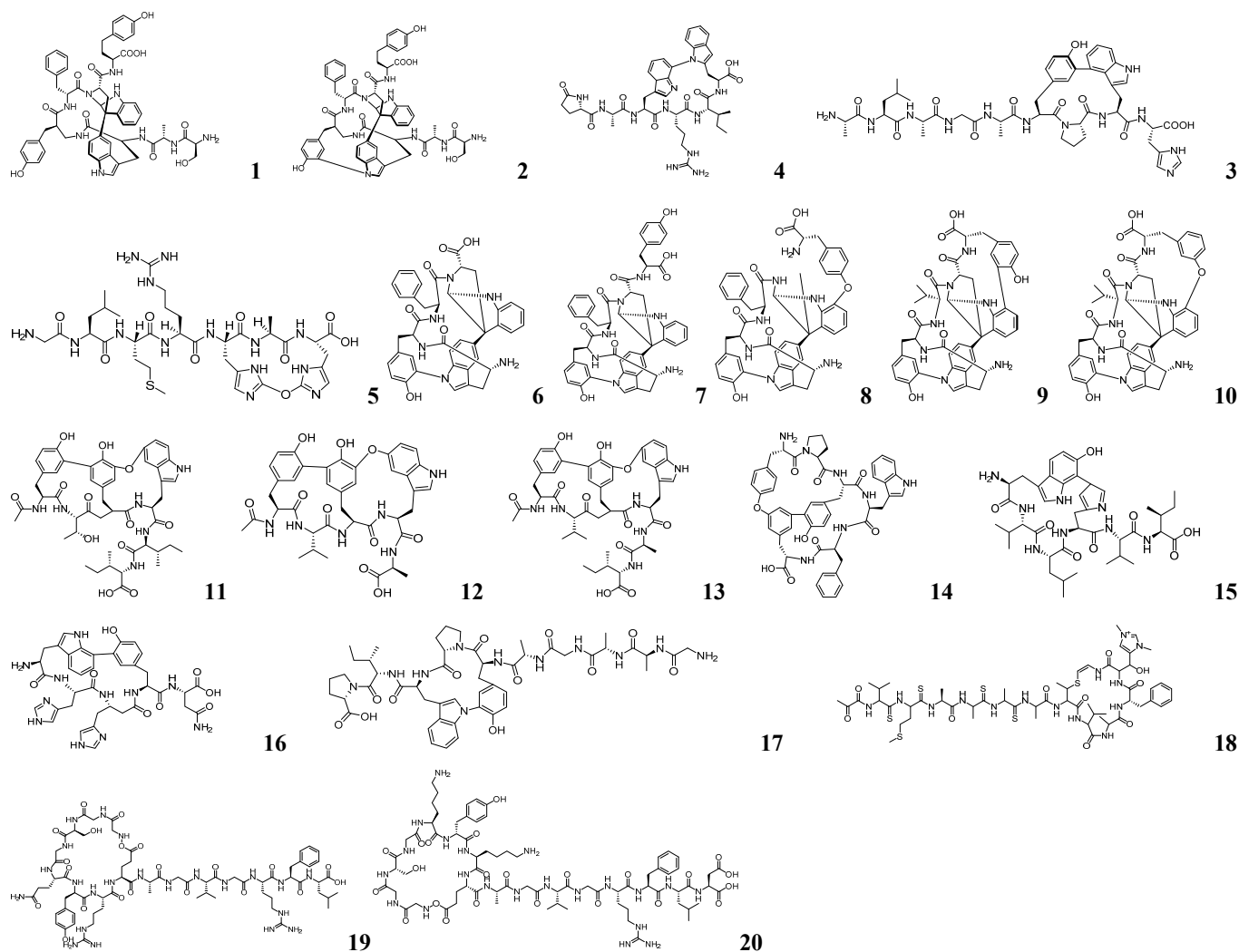


Figure 2. The chemical structures of RiPPs by genome mining (1–20).

Position-Specific Iterated Basic Local Alignment Search Tool (PSI-BLAST) detects tele-evolutionary relationships between proteins, identifies conserved structural domains or motifs in protein sequences, and optimizes the search results by iteratively constructing a position-specific scoring matrix [66]. Rapid ORF Description and Evaluation Online (RODEO) is a bioinformatics tool designed for the rapid identification and evaluation of open reading frames (ORFs) [67]. It is particularly well-suited for predicting potentially functional ORFs from metagenomic or transcriptomic data. Seokhee Kim and his team utilized these two tools to establish the connection between P450 and RiPPs [68]. The approximate mining process is shown in Figure 3A. Using TrpB (associated with aromatic-containing peptides) and BytO (linked to coaromatic peptides) as query sequences, researchers employed PSI-BLAST to identify evolutionarily related P450 homologs, while RODEO facilitated efficient prediction of adjacent ORFs encoding potential precursor peptides with characteristic aromatic features. This dual-tool strategy enabled the identification of 19 novel BGCs through PSI-BLAST analysis. Three of the precursor peptides predicted by these BGCs have rare aromatic amino acid residues and remain well-translated when co-expressed with the corresponding P450 enzymes. Following heterologous expression of the BGCs corresponding to the three selected precursor peptides in *E. coli*, three new

modified peptides, namely roseovertin (15), rubrin (16) and lapparbin (17) were obtained, which are shown in Figure 2. These findings led to the establishment of “cyptides” as a new RiPP superfamily, defined by their characteristic P450-mediated multi-linkage aryl cross-bridges (Trp C7'-to-His N- τ , Tyr C-6-to-Trp N-1', and Trp C-7'-to-Tyr C-6), which substantially expands the structural diversity of post-translationally modified peptides. This method demonstrates how strategic integration of PSI-BLAST's sensitive homology detection with RODEO's efficient ORF prediction can effectively make a connection between P450 enzymes and their cognate RiPP products, providing a robust framework for future discovery of complex peptide modifications.

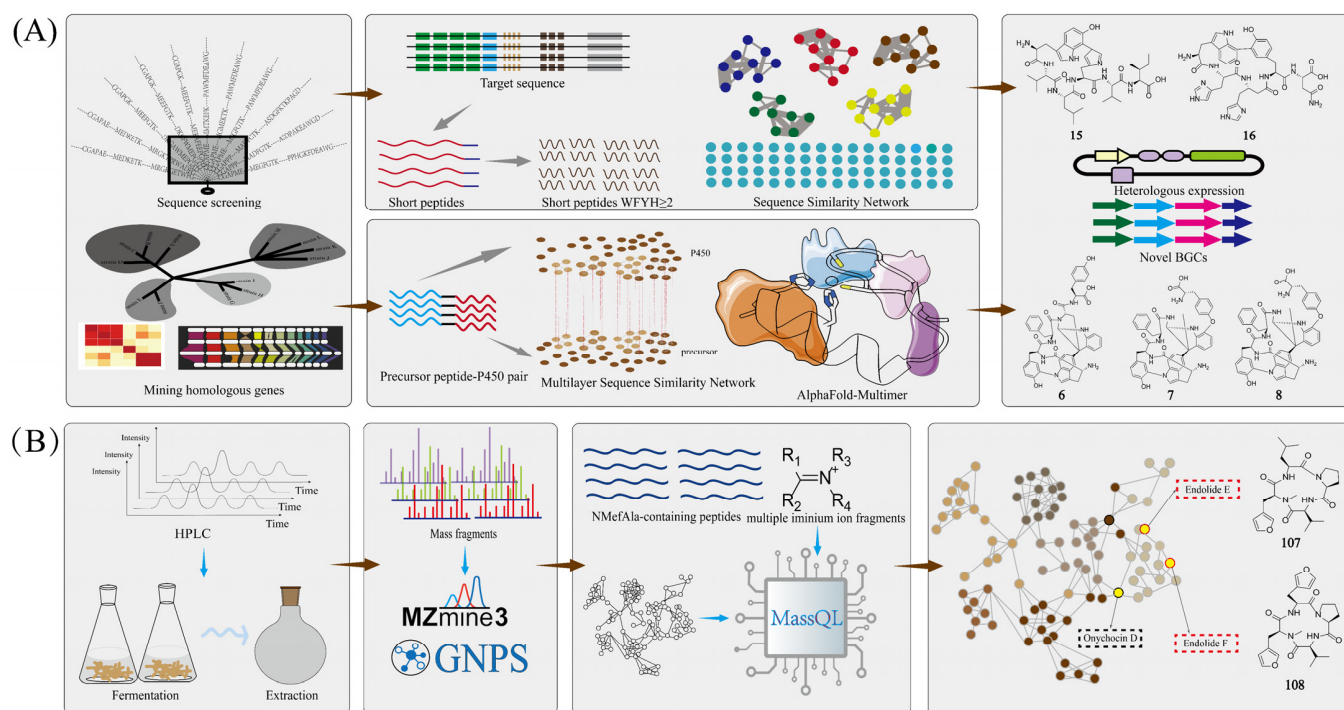


Figure 3. (A) Examples of genome mining for new NPs. (B) Examples of metabolomic mining for new NPs.

The integration of diverse bioinformatics tools facilitated the establishment of precise correlations between P450 enzymes and RiPPs, enabling the systematic mining of such RiPPs. Screening for the genes with high similarity or homologs as target genes is an efficient approach for identifying other members of a compound family. The genes identified through screening often fail to express the target product under laboratory culture conditions and typically require heterologous expression or genetic modification [51,69,70].

The λ -Red system, derived from the λ -Red phage, overcomes the limitation of restriction endonuclease cleavage sites through homologous recombination technology [71,72]. It is widely used for various genetic modifications and is one of the most common tools in gene editing. Kazuo Shin-ya and his team employed BLAST analysis to identify a precursor peptide gene containing the VMAAAATVAFHC motif in *Streptomyces* sp. MSB090213SC12, showing high similarity to the known thioredoxin sequence (VMAAAASIALHC) within a genome database comprising over 100 actinomycetes [73]. The target neothiocyanin biosynthetic gene cluster (*ntv*) was isolated through the construction of a Sau3AI-digested BAC library and subsequent PCR screening, yielding the clone pKU503J143P1-J10. Minimized gene clusters were obtained following modification of the cloned genes using the λ -Red system. These optimized clusters were heterologously expressed in *S. avermitilis* SUKA22, leading to the production of the novel thioredoxin compound neothioviridamide (18), which is shown in Figure 2. The approach demonstrates how combining computational pre-

diction with advanced genetic manipulation can unlock the biosynthetic potential of silent gene clusters that remain refractory to conventional fermentation-based discovery methods.

Genetic annotation of the fungal-bacterial endosymbiotic system *Mycetohabitans rhi-zoxinica*, conducted by Christian Hertweck and his team, revealed its potential to produce burhizin and mycetohabin-like lasso peptides [73]. Initial heterologous expression attempts in *E. coli* BL21 successfully yielded mycetohabin-15 (19) and mycetohabin-16 (20), while the burhizin-23 (21) construct (vector pEB45) produced only a truncated variant due to expression incompatibility. The structures of these three compounds as shown in Figures 2 and 3. Considering host compatibility, the mushroom pathogen *Burkholderia gladioli* pv. *agaricola* HKI0676 was selected as a heterologous expression host, which successfully generated the target compound. This study not only expanded the known structural diversity of lasso peptides but also established a critical precedent for host selection in heterologous expression systems, demonstrating that heterologous expression hosts can overcome expression barriers encountered in conventional bacterial platforms.

Shang-Wen Luo and his team analyzed the genome of *Streptomyces yunnanensis* using antiSMASH, successfully identifying two linear bacteriocin BGCs, named *yan* and *yn* [74]. During heterologous expression, *yan* and *yn* were individually cloned into the *E. coli*-*Streptomyces* shuttle vector pSET-kasO which contains a strong promoter, kasO, upstream of the cloning site to drive exogenous gene expression. Screening of different host strains and culture media ultimately enabled the expression of the *yan* in *Streptomyces azureus* M1154 and the *yn* in *Streptomyces variegatus* GX28. This successfully resulted in the isolation of three linear anti-mycobacterial compounds, yunnanaridins A–C (22–24), which are shown in Figure 4. This approach demonstrated the critical importance of matching specific biosynthetic pathways with compatible expression hosts, as evidenced by the differential expression outcomes between the two gene clusters.

The Ion Torrent Personal Genome Machine (PGM) is a sequencing platform that operates on the principle of detecting changes in hydrogen ion (H^+) concentration released during DNA synthesis, enabling high-throughput sequencing [75]. Rapid Annotation using Subsystem Technology (RAST) is an automated microbial genome annotation platform that utilizes the subsystem functional classification framework [76]. It facilitates gene function prediction through comparisons with known functional protein databases. Juan A. Asenjo and his team employed an integrated genomic approach to characterize the novel actinomycete *Streptomyces huasconensis* HST28T, combining Ion Torrent PGM with RAST [77]. This strategy enabled the identification of a compact lasso peptide biosynthetic gene cluster (*hpt*), whose heterologous expression produced huascopeptin (25), a novel Gly1-Asp7 macrocyclic lactam ring lasso peptide, which is shown in Figure 4. Remarkably, huascopeptin (25) represents the smallest known lasso peptide to date, featuring an unusually compact ring-tail architecture that defies conventional structural paradigms for this class of ribosomally synthesized peptides. This discovery not only expands the known size range and structural diversity of lasso peptides but also provides new insights into their structure–activity relationships.

2.2. Genome Mining of Terpenoids

Terpenoid genome mining employs distinct strategies for fungal and bacterial systems, focusing on key enzymatic signatures in their respective biosynthetic pathways [78]. The primary targets for terpenoid genome mining in fungi are typically individual TSs and prenyltransferase terpene synthases (PTTs) [79]. The catalytic core of PTTs comprises two or more independent terpene synthase modules, which mediate multi-step cyclization reactions to generate structurally complex hybrid terpenoids [80,81]. Bacterial terpenoid

mining, conversely, emphasizes the identification of specialized post-modification enzymes that impart structural diversity to core terpene scaffolds.

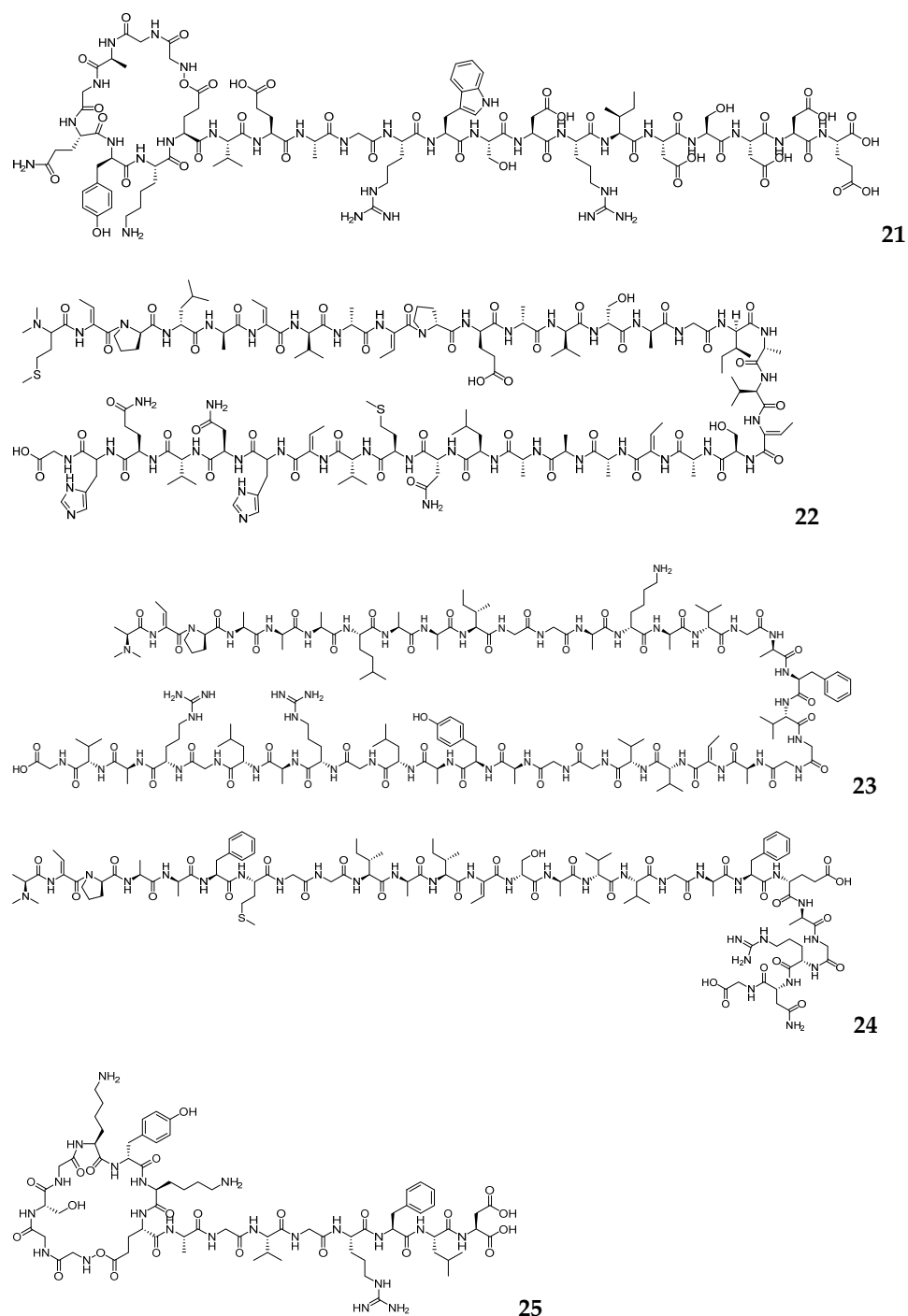


Figure 4. The chemical structures of RiPPs by genome mining (21–25).

HMMER is a widely used suite of tools for biological sequence analysis based on HMMs, enabling protein/nucleic acid sequence homology searches, gene family classification, and functional annotation [82,83]. Tian-gang Liu and his team developed an efficient workflow for discovering novel PTTS [80]. Firstly, using HMMER comprehensive searches across 519 fungal genomes in the Pfam database to identify sequences containing both PF03936 (prenyltransferase) and PF00348 (terpene synthase) domains—the characteristic signature of PTTS enzymes. Secondly, the team screened the NCBI and UniProt databases for PTTS genes, selecting hypothetical TSs with an amino acid sequence length of ap-

proximately 700 and conserved motifs within the PT and TS structural domains [84–87]. Combining these two datasets and retaining PTTS with less than 80% similarity, 74 candidate PTTS genes were identified. Functional characterization of the 74 candidate PTTS genes was performed using a high-efficiency precursor-providing yeast chassis coupled with a high-throughput automated platform. This system enabled rapid construction and analysis of numerous yeast mutants, revealing 34 functional PTTS enzymes including PTTS010 which produces sesterevisene (**26**), and PTTS009/037/054 which generates sestororbiculene (**27**). The structures of these two compounds are shown in Figure 5. This study significantly accelerated the functional characterization of PTTSs by integrating a high-efficiency precursor-providing yeast chassis with a high-throughput automated platform, enabling the rapid construction of numerous yeast mutants from gene fragments. This versatile system is also applicable to the functional study of other terpene synthases.

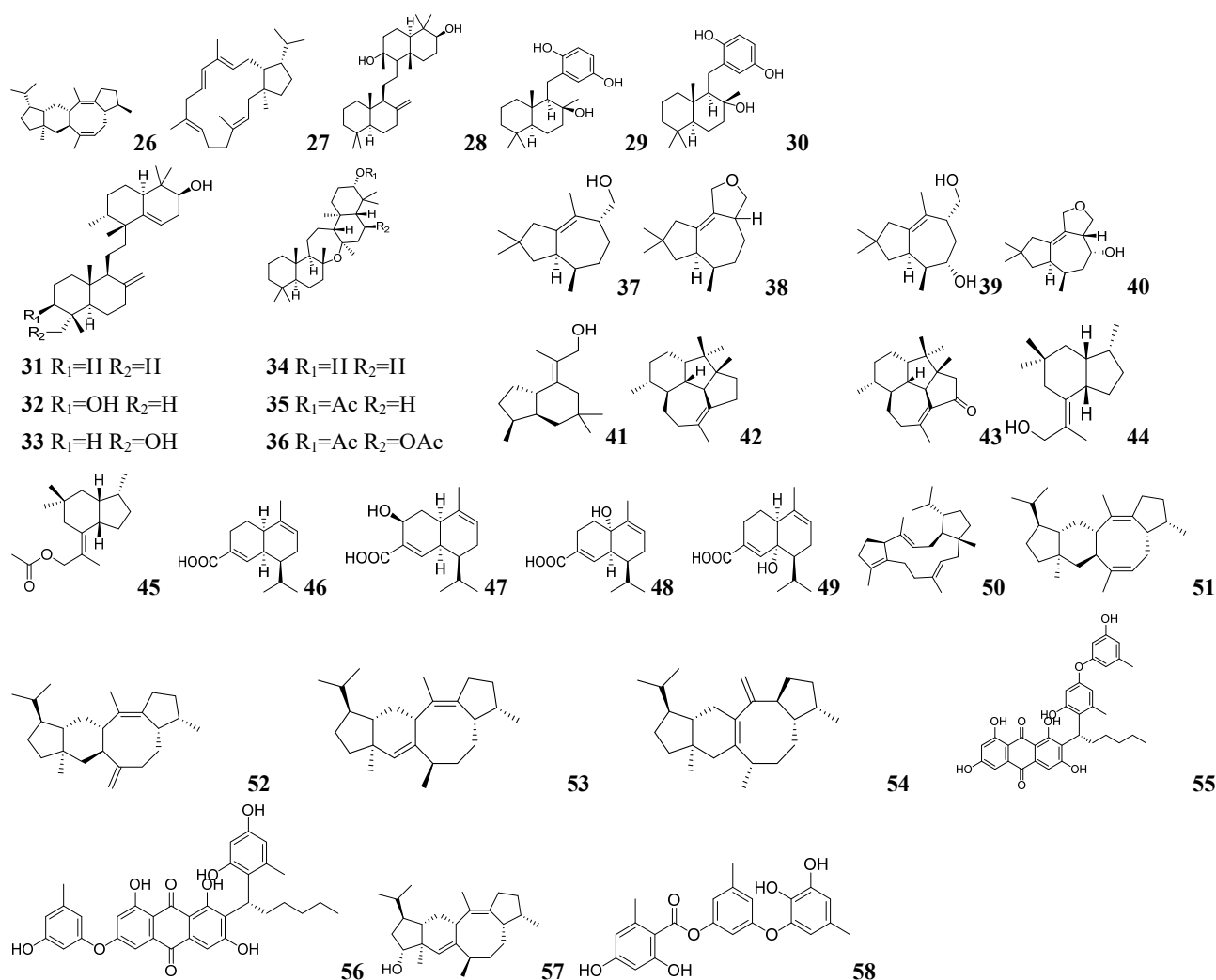


Figure 5. The chemical structures of terpenoids by genome mining (**26**–**58**).

Yudai Matsuda and his team developed FunBGCeX, an innovative fungal genome mining tool that addresses the challenge of identifying BGCs associated with domain-free enzymes [88]. The team first constructed a fungal BGCs (FunBGCs) database containing approximately 700 fungal BGCs, from which they extracted 5070 protein sequences for comprehensive Pfam domain analysis using HMMER. This analysis revealed 572 proteins lacking conventional domain architecture, highlighting a significant gap in current mining approaches. To overcome this limitation, the team developed specialized hidden Markov models (HMMs) targeting PKS and NRPS domains and constructed a DIAMOND database

incorporating all biosynthesis-related proteins. Screening of BGCs encoding terpene cyclase Pyr4 homologs from 1990 fungal reference genomes was performed using FunBGCeX. Three BGCs (*homo*, *fumi*, and *alli*) encoding SHC/OSC-like proteins from different fungi and one BGC (*mos*) from *Neoarthriniium moseri* CBS 164.80 were identified through screening [89]. Nine new terpenoids (28–36) were produced by cloning the relevant genes into expression vectors and heterologously expressing them in *Aspergillus oryzae* NSAR1 or NSARU1. The structures of these compounds are shown in Figure 5. Unlike conventional mining software, FunBGCeX combines manually curated reference data with customized HMM profiling to specifically target atypical, domain-free enzyme-associated BGCs with superior precision and efficiency, establishing a new paradigm for uncovering fungal metabolic diversity.

There are many more studies of fungal TSs. After whole genome sequencing of *Irpex lacteus* by Guang-kai Bian and his team, gene cluster annotation was performed using antiSMASH [90]. Their integrated bioinformatics approach identified 19 potential sesquiterpene synthase (STS) genes, from which sequence comparison revealed 10 non-alleles. Heterologous expression of the selected *Il4946* gene in *Aspergillus oryzae* produced five new sesquiterpenoids: iltremulanols A–D (37–40) and trichobrasilenol (41), which are shown in Figure 5. Sheng-ying Li and his team identified a terpene BGC (*ven*) in *Streptomyces venezuelae* ATCC 15439 through gene annotation [91]. This BGC encodes the production of two diterpene compounds, namely venezuelaenes A–B (42–43), as shown in Figure 5, which feature a novel 5-5-6-7 tetracyclic backbone. Using known PaFS and AcOS sequences as probes, Jaclyn M. Winter and his team identified a cluster of potential bifunctional TS genes (*tnd*) in the genome of *Aspergillus flavipes* CNL338 [92]. This gene cluster encodes the production of a novel diterpene, namely talarodiene, which features a unique tricyclic structure. Tian-gang Liu and his team sequenced and annotated the genome of *Trichoderma viride* J1-030, identifying the potential TS gene *Tvi09626* through screening [93]. The gene catalyzed the production of a novel five or six bicyclic sesquiterpene (44) and its esterified derivatives (45), which are shown in Figure 5. This is the first sesquiterpene synthase characterized in this fungus, enriching the diversity of terpenoids in *Trichoderma viride*.

CD-HIT (Cluster Database at High Identity with Tolerance) is a widely used bioinformatics tool for clustering and comparing protein or nucleotide sequences [94,95]. BiG-SCAPE (biosynthetic gene similarity clustering and prospecting engine) is a bioinformatics tool designed for analyzing and comparing BGCs [96]. It performs cluster analysis based on sequence similarity and gene function to identify groups of genes with similar biosynthetic potential. Pei-Yuan Qian and his team utilized CD-HIT and BiG-SCAPE to comprehensively screen NCBI bacterial genomes and identified 2892 cytochrome P450-containing terpene synthase/cyclase genes that were organized into 355 distinct cluster families [97]. Focusing on the phylogenetically unique *lev* gene cluster that formed a discrete group in SSN, were selected for heterologous expression in *Streptomyces azureus* M1146. This led to the production of four novel α -amorphene type sesquiterpenes, namely levinoids A–D (46–49), as shown in Figure 5. Similarly, Kui Hong and his team identified the candidate gene *Au11189* from the *Aspergillus ustus* 094102 genome through BLAST search and functional prediction, naming it *AuAS* [98]. Initial heterologous expression of *AuAS* in *Aspergillus oryzae* NSAR1 produced five novel sesquiterpenes, namely aspergiltriene A (50) and aspergildienes A–D (51–54), while subsequent co-expression with the adjacent cytochrome P450 gene *AuAP450* yielded four structurally distinct dibenzoterpenoid alcohols aspergilols A–D (55–58). The structures of these compounds are shown in Figure 5. This method focuses on genome mining of modifying enzymes, overcoming the limitation of relying solely on TSs for novel compound discovery. It provides a more efficient approach to identify new bacterial terpenoids.

2.3. Genome Mining of Polyketides

Seq2PKS is a machine learning-based algorithmic tool developed by Hosein Mohimani and his team for predicting the chemical structures of type I *cis*-acyltransferase (AT) polyketide compounds [99]. Its specific functions include the following: (1) predicting compounds produced by type I PKS gene clusters and improving prediction accuracy through mass spectrometry database calibration; (2) predicting acceptor substrate ranges of AT domains using an ExtraTrees-based classification algorithm, achieving an overall accuracy of 94%; (3) predicting the biosynthetic logic of *cis*-AT polyketides using a paired nearest neighbour (PNN)-based search combined with gene order in the genome [100]; (4) predicting polyketide core structure post-modification using a custom enzymatic modification database. When applied to analyze *Streptomyces actiphen* genomes from the NRRL collection, Seq2PKS identified a novel long-chain polyketide system with 55% similarity to known actiphenol biosynthetic pathways generating 258 structural predictions. Using the in-house Dereplicator+ tool, the team identified two mass spectrometry matches to polyketides, including the novel compound 2-aminobenzamide-actiphenol (**59**) upon expression, as shown in Figure 6 [101]. The algorithm integrates mass spectrometry data to predict structures of diverse mature polyketides, including those with unknown modifications. It outperforms PRISM and antiSMASH in benchmarking for *cis*-AT polyketide BGC prediction.

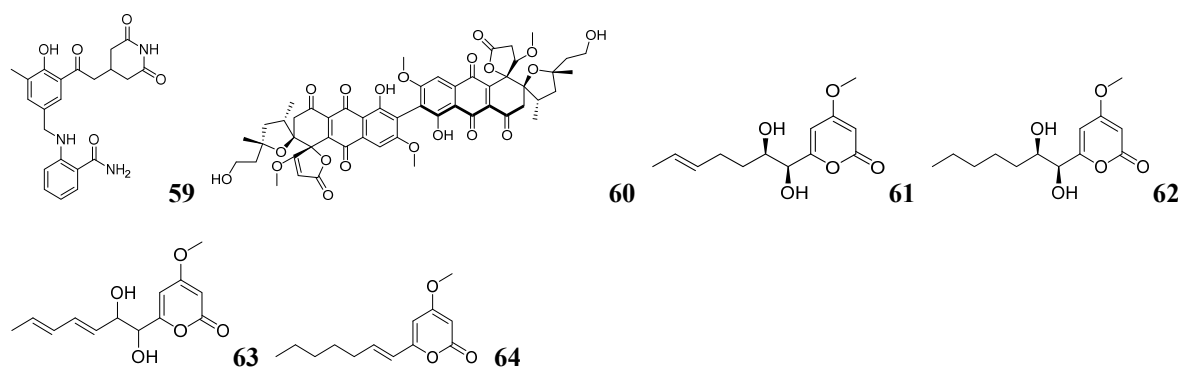


Figure 6. The chemical structures of polyketides by genome mining (**59–64**).

Hui-Ming Ge and his team screened 21,728 actinomycete genomes (NCBI database and in-house sequencing data) via BLAST 2.14.0, identifying 5944 type II PKS and 919 KAS III-containing gene clusters [102]. Two homologous gene clusters (*spi* and *msp*) from *Streptomyces spectabilis* NA07643 and *Micromonospora* sp. HM134 were finally identified. AntiSMASH annotation of the *spi* gene cluster revealed genes encoding type I PKS, type II PKS, KAS III, and multiple modifying enzymes, enabling the successful heterologous expression of spirocycline A (**60**) as shown in Figure 6. Spirocycline A (**60**) showed significant inhibition against *Micrococcus garciniae* (MIC = 2 µg/mL). Genome sequencing of the endophytic fungus *Calcarisporium arbuscula* by Ling Liu and his team resulted in the identification of 68 potential BGCs, including a highly reduced polyketide synthase (HRPKS) cluster *cpn* [103]. Heterologous expression of constructed plasmids containing *cpnA*, *cpnB*, and *cpnC* in *Aspergillus oryzae* A1145 yielded the novel α-pyranones calcapyrones A–B (**61–62**), along with biosynthetic intermediates calcapyrones C–D (**63–64**), as shown in Figure 6. Calcapyrone C (**63**) is weakly cytotoxic to cancer cells.

3. Metabolome Mining NPs

3.1. Metabolome Mining Based on MS

Ultra-performance liquid chromatography coupled with quadrupole-time-of-flight tandem mass spectrometry (UPLC-Q-TOF-MS/MS) is a high-resolution analytical technique

for accurate compound identification and quantification [104]. UPLC significantly reduces analysis time and improves sensitivity. The Q-TOF system fragments ions via CID and screens target ions by interpretation of m/z patterns to infer structural features [105]. UPLC-Q-TOF-MS/MS is the leading technique for metabolomics, unknown compound identification, and precise quantification, combining ultrafast separation with high-resolution mass detection. Gang Ding and his team used UPLC-Q-TOF-MS/MS to screen for compounds containing three characteristic fragments when mining secondary metabolites from *Chaetomium cochliodes* [106]. Previous studies confirmed that chetomin analogs with polysulfide bridges are cyclic peptides exhibiting an asymmetric dimeric structure. The key biosynthetic step in chetomin analog production is dimerization, which generates numerous characteristic fragment ions. The mass spectra of chetocochliodins C/D/E reveal shared characteristic fragments at m/z 270, m/z 284, and m/z 282. Structural analysis revealed these fragments derive from either (1) $\text{HOCH}_3/\text{HSCH}_3$ elimination from protonated molecules or (2) sequential loss of two $\text{HOCH}_3/\text{HSCH}_3$ units, enabling classification into three structural subtypes (A–C). Using UPLC-Q-TOF-MS/MS to target compounds containing the three characteristic fragments, the team successfully identified two novel chetomin analogs, namely chetocochliodins M (65) and N (66), as shown in Figure 7. Chetocochliodin N (66) demonstrates potent cytotoxicity against A549 and HeLa cancer cell lines, exhibiting superior activity compared to the positive control cisplatin. These findings highlight its potential as a lead compound for anticancer drug development. The results further highlight the crucial role of sulfur bridges in mediating the cytotoxicity of chetomin analogs, providing a foundation for investigating their structure–activity relationships.

GNPS is an open-access platform that utilizes MS data for molecular network analysis and natural product annotation. GNPS has three main advantages: (1) free and open source, which reduces the cost of scientific research; (2) a powerful public database with wide coverage of compound types; (3) richer data shared by scientists worldwide. The GNPS workflow is straightforward: users simply convert LC-MS/MS data to .mzXML format and upload the files to the platform (<https://gnps.ucsd.edu>, accessed on 15 March 2025), which automatically generates interactive molecular networks for visualization and analysis. GNPS is a powerful tool in metabolomics with a wide range of applications. Jun-Feng Wang and his team employed a targeted metabolomics approach to discover novel 4-hydroxy-2-pyridone alkaloids from sponge-derived *Arthrinium arundinis* ZSDS1-F3 [107]. By integrating LC-MS/MS analysis with GNPS molecular networking, they identified a characteristic cluster containing known arthpyrone alkaloids C and G, which guided the isolation of three structurally novel analogs—arthpyrones M–O (67–69) as shown in Figure 7. Arthpyrones M–O (67–69) demonstrated cytotoxic activity against some cancer cell lines ($\text{IC}_{50} = 0.26\text{--}6.43\ \mu\text{M}$). Arthpyrone O (69) had strong inhibitory and pro-apoptotic effects on SCLC cell lines in vitro and significantly inhibited SCLC cell xenograft tumor growth in in vivo experiments. This study identifies promising new drug candidates for SCLC therapy while expanding the potential anticancer applications of 4-hydroxy-2-pyridone alkaloids.

Yong-Hong Liu and his team employed the GNPS platform to explore tanzawaic acid (TA) derivatives from *Penicillium steckii* SCSIO 41025 [108]. They successfully isolated and identified 40 TA derivatives, including 22 novel compounds (70–91), as shown in Figure 7. Most of these compounds inhibited LPS-induced NF- κ B activity and RANKL-induced osteoclast differentiation. Christine Beemelmans and his team employed LC-MS coupled with GNPS-based metabolomics to isolate four novel ergosterol derivatives, namely podaxisterols A–D (92–95), as shown in Figure 7 from *Podaxis* sp. Ethiopia [109]. During metabolomic analysis of *Pseudomonas* spp. strain FhG100052, Jens Glaeser and his team identified a dense cluster of cyclic lipopeptides (CLPs) exhibiting rich molecular

networking features [110]. Through optimization of culture conditions, they successfully isolated five novel CLPs: stechlisin B2 (96), stechlisin F (97), tensin (98), stechlisin D3 (99), and stechlisin C3 (100). The structures of these compounds are shown in Figures 7 and 8. The antimicrobial activity of stechlisins against *Moraxella catarrhalis* FH6810 exhibited a strong correlation with lipid chain length. Specifically, stechlisin B2 (96) had no inhibitory effect, tensin (98) had moderate activity, and stechlisin F (97) had significant activity. Christine Bemelman and his team discovered that *Amycolatopsis saalfeldensis* inhibited *Pseudoxylaria* sp.X802 in co-culture [111]. GNPS analysis of large-scale cultures yielded three novel type II thiopeptides, namely saalfelduracins B–D (101–103), as shown in Figure 8, which showed antimicrobial activity against Gram-positive bacteria.

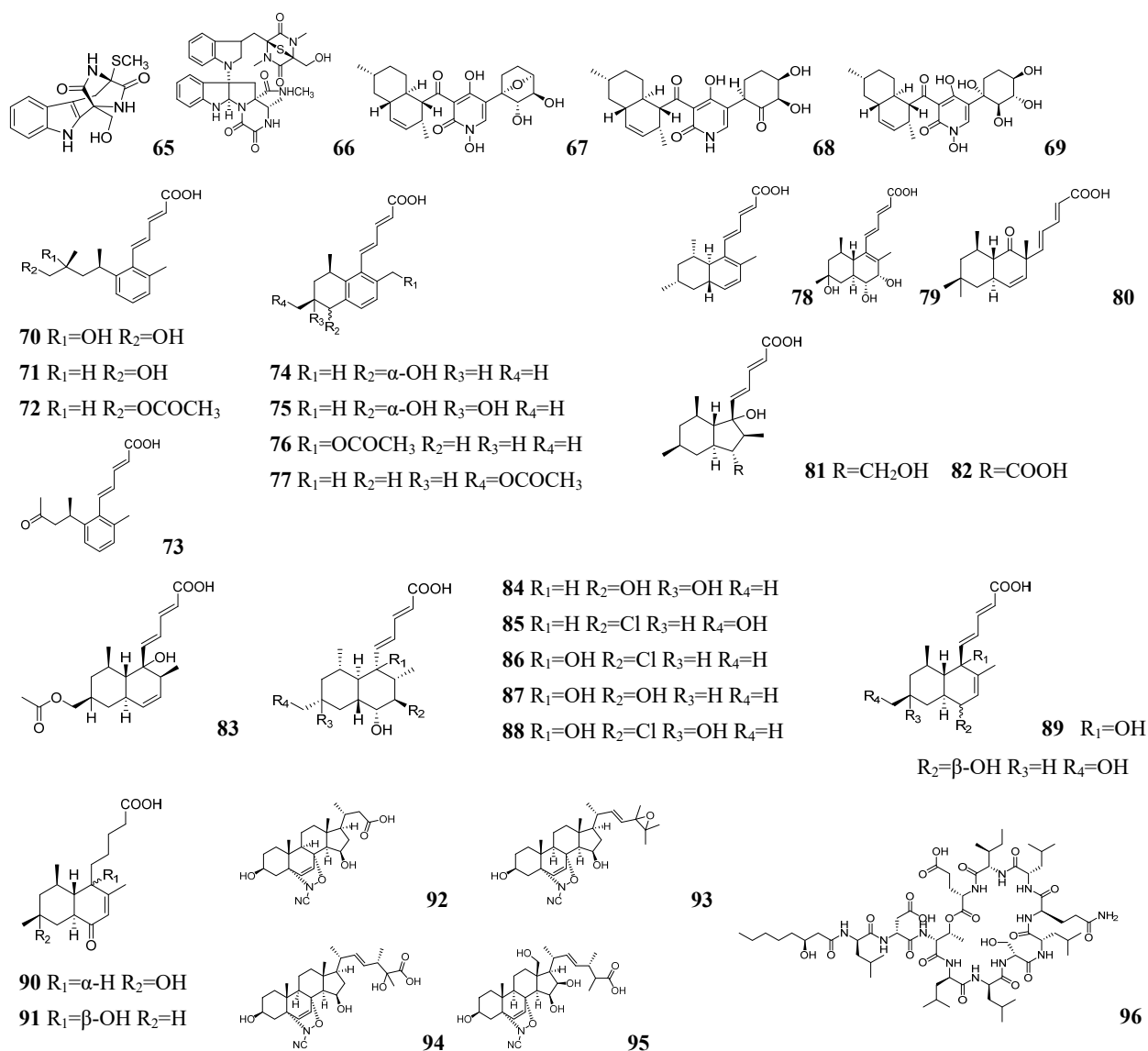


Figure 7. The chemical structures of NPs by metabolome mining based on MS (65–96).

FBMN is an advanced LC-MS/MS analysis method that enhances molecular network accuracy and quantification by integrating MS¹ feature peaks with MS² spectra. Compared with traditional GNPS, FBMN has four main advantages: (1) it avoids erroneous clustering due to retention time drift in traditional GNPS through RT correction and peak alignment and can distinguish isomers; (2) it correlates MS¹ peak intensities, which allows cross-sample comparisons such as metabolite differences in different culture conditions; (3) unlike traditional GNPS that mainly relies on DDA (data dependent acquisition), FBMN

can handle DIA (data non-dependent acquisition) to improve the detection rate of low abundance compounds [112,113]; (4) FBMN can integrate tools such as SIRIUS for more accurate structure prediction. In their investigation of secondary metabolites from *Epicoccum* sp. 1-042, Yun-Ying Xie and his team performed LC-MS/MS analysis of fractionated extracts [114]. The MS data were processed using MZmine, analyzed via FBMN on the GNPS platform, and visualized with Cytoscape software (v3.10.3). Polycyclic tetrahydropyrimidinic acids containing cis-decahydronaphthalene, namely epicolidines A–C (104–106), as shown in Figure 8 were isolated and identified, of which epicolidine B (105) and C (106) exhibited inhibitory activity against Gram-positive bacteria. In recent years, FBMN has evolved from primarily serving as a compound discovery tool to becoming increasingly valuable for metabolomic data processing and target compound prioritization.

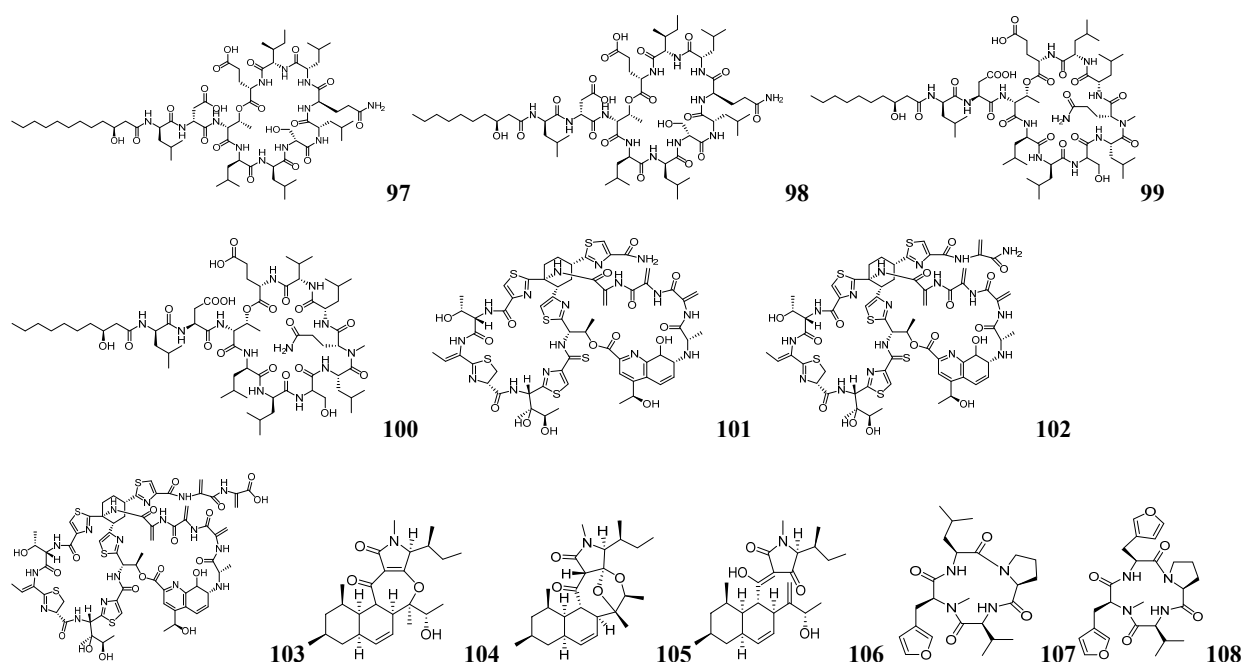


Figure 8. The chemical structures of NPs by metabolome mining based on MS (97–108).

Mass Spectrometry Query Language (MassQL) enables rapid extraction of target information from complex LC-MS/MS data [115]. The method enables automated, high-precision screening of target compounds through customizable parameters including parent ions, fragment ions, and neutral loss patterns. This automated analytical process effectively replaces traditional manual peak extraction and facilitates efficient batch processing of large datasets. Raphael Reher and his team developed an innovative mass spectrometry workflow combining MassQL with customized computational scripts to systematically identify compounds containing the diagnostic N-methyl-3-(3-furoyl)-alanine (NMefAla) fragment [116]. The approximate mining process is shown in Figure 3B. They employed MassQL's targeted query capabilities to screen LC-MS/MS datasets for diagnostic fragment patterns, complemented by two specialized scripts—the first identifying NMefAla-containing structures and the second applying iminium ion filters to reduce false positives. Following GNPS-FBMN molecular networking, MassQL-enabled substructure annotation of network nodes facilitated the discovery of two novel proline-derived endocannabinoids, namely endolides E–F (107–108), as shown in Figure 8. Endolide F (108) has moderate antagonistic activity against arginine pressin V_{1A} receptors. While MassQL-annotated molecular networks enhance structural characterization by adding substructural information to facilitate novel compound identification, the approach presents several

limitations: (1) demanding instrumentation requirements, and (2) inability to perform MS² signal quantification.

3.2. Metabolome Mining Based on NMR

¹H NMR is a powerful analytical technique with powerful capabilities for the structural identification of organic compounds, mixture composition analysis, and isomer differentiation. ¹H NMR chemical shifts exhibit characteristic ranges that reflect distinct structural features: aliphatic protons (R-CH₃): δ_H 0.5–2.0 ppm; olefinic protons (=C-H): δ_H 4.5–6.5 ppm; aromatic protons (Ar-H): δ_H 6.5–8.5 ppm; carboxylic acid protons (-COOH): δ_H 10–13 ppm. Hang-Lun Shao and his team pioneered an integrated analytical approach combining HPLC-MS/MS molecular networking with ¹H NMR spectroscopy to discover novel bioactive peptides from 270 marine-derived *Penicillium* species [117]. Their workflow initially identified a distinctive metabolite cluster through characteristic neutral amino acid fragment losses in mass spectra, with subsequent ¹H NMR analysis of target extracts revealing diagnostic amino acid signatures (exchangeable protons at δ_H 9.00–7.50 and α -protons at δ_H 4.50–3.50). Seven new peptide analogs chrysogeamides A–G (109–115) as shown in Figure 9 were isolated. The study establishes a robust template for accelerating the identification of bioactive NPs from complex microbial extracts by coupling molecular networking with NMR.

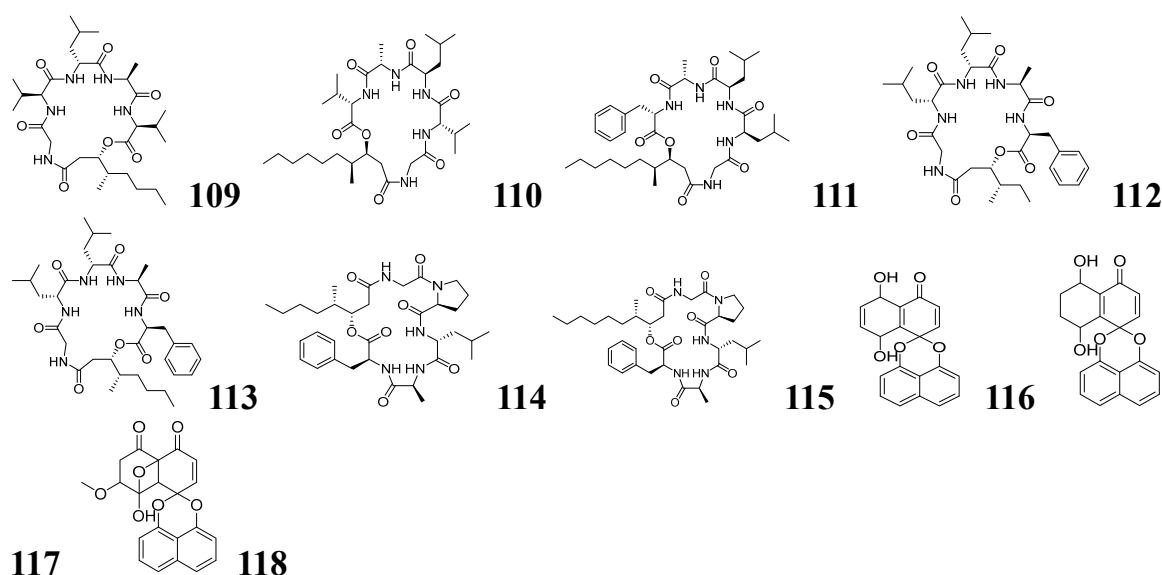


Figure 9. The chemical structures of NPs by metabolome mining based on NMR (109–118).

MADByTE is a mixture analysis method based on NMR and diffusion-ordered spectroscopy (DOSY), which is mainly used for the separation and component identification of complex mixtures [118–120]. The method calculates diffusion coefficients based on differential diffusion rates of molecules with varying sizes in solution, separates mixture components into distinct diffusion groups according to these coefficients, and subsequently identifies molecular structures within each group by integrating chemical shifts with coupling constant data. MADByTE enables component analysis of complex mixtures without physical separation. Nicholas H. Oberlies and his team used the MADByTE platform for dereplication in mining fungal metabolites for the resorcylic acid lactones (RALs) and spirobisnaphthalenes with good results [118]. By establishing a reference database containing heteronuclear single quantum coherence (HSQC) and total correlation spectroscopy (TOCSY) spectra of 19 RALs and 10 spirobisnaphthalenes, they enabled efficient on-target/off-target discrimination through MADByTE analysis. After investigating the

structural features and bioactivities of RALs and spirobisnaphthalenes, an analysis of the full association network using the MADByTE platform showed that RALs with specific structures formed clusters in the bioactivity network, revealing a structure–activity relationship. The data from seven fungal extracts were analyzed in the above process, identifying strain MSX64790 as the target. Through combined mining using the MADByTE platform and 2D spectral analysis, three spirobisnaphthalene-like compounds palmarumycins CP20–CP22 (**116–118**) as shown in Figure 9 were obtained. This study not only marks the first successful implementation of MADByTE for fungal metabolite dereplication but also demonstrates its unique capability to visually map bioactivity patterns onto structural networks, establishing a powerful new paradigm for conducting SAR studies directly in complex mixtures.

4. Genomic and Metabolomic Guided the Isolation of NPs

4.1. Genomics Combined with Isotope Labeling and NMR

The landscape of bioinformatics tools for microbial natural product discovery features complementary platforms like antiSMASH and RODEO for BGC analysis, with antiSMASH 5.1 specifically enhancing RiPPs identification through improved cluster boundary prediction and recognition algorithms compared to its predecessors. Parallel developments include RIPPMiner, which adopts distinct computational approaches from RODEO while serving similar functions in RiPP-associated BGC prediction [121,122]. Isotope feeding experiments are a classical method for tracing metabolic pathways, elucidating biosynthetic mechanisms, and studying biochemical transformations using stable or radioactive isotope-labeled precursors [123]. Isotope feeding experiments combined with NMR and the above bioinformatics tools can efficiently resolve the synthesis logic and backbone structure of metabolites.

Harald Gross and his team developed an integrated approach combining bioinformatics prediction with advanced NMR-guided isotope tracing to discover novel RiPPs in *Nocardia terpenica* strains IFM 0406 and IFM 0706T [124]. Using antiSMASH 5.1, RODEO, and RIPPMiner, they identified the *nta* gene cluster encoding putative thiazole/oxazole-modified RiPPs. To characterize the metabolites encoded by the *nta* gene cluster, isotope feeding experiments were performed using $(^{15}\text{NH}_4)_2\text{SO}_4$, $[^2\text{H}_{10}]$ -L-leucine, and $[^2\text{H}_7]$ -L-proline. The results indicated that the target compound is a 12- to 13-amino acid polypeptide. Structural analysis, combined with genomic data on characteristic enzymes, revealed the presence of thiazole/oxazole moieties. Targeted screening was performed for compounds with molecular weights >1300 Da exhibiting characteristic ^1H - ^{13}C HMBC correlations of oxazole (δ_{H} 8.18/ δ_{C} 148.32) or thiazole (δ_{H} 8.18/ δ_{C} 174.91) moieties. Three new RiPPs were successfully isolated from *N. terpenica* IFM 0406 and 0706 and named nocathioamides A–C (**119–121**) as shown in Figure 10. This study utilized bioinformatics tools to predict the structural features of nocathioamides. By integrating stable isotope labeling with ^1H - ^{13}C -HMBC analysis successfully identified the target peptides from complex metabolite mixtures, thereby resolving the key challenge of linking biosynthetic genes to their metabolic products in RiPP research. This approach is applicable not only to thiazole-containing RiPPs in *Nocardia*, but can also be adapted to discover other RiPPs with rare functional groups. The integration of ^1H - ^{15}N -HSQC screening allows the study of imine-containing RiPPs; the adjustment of ^1H - ^{13}C -HMBC sequences can tap into more thioamidated RiPPs, which provides a general idea for the study of RiPPs.

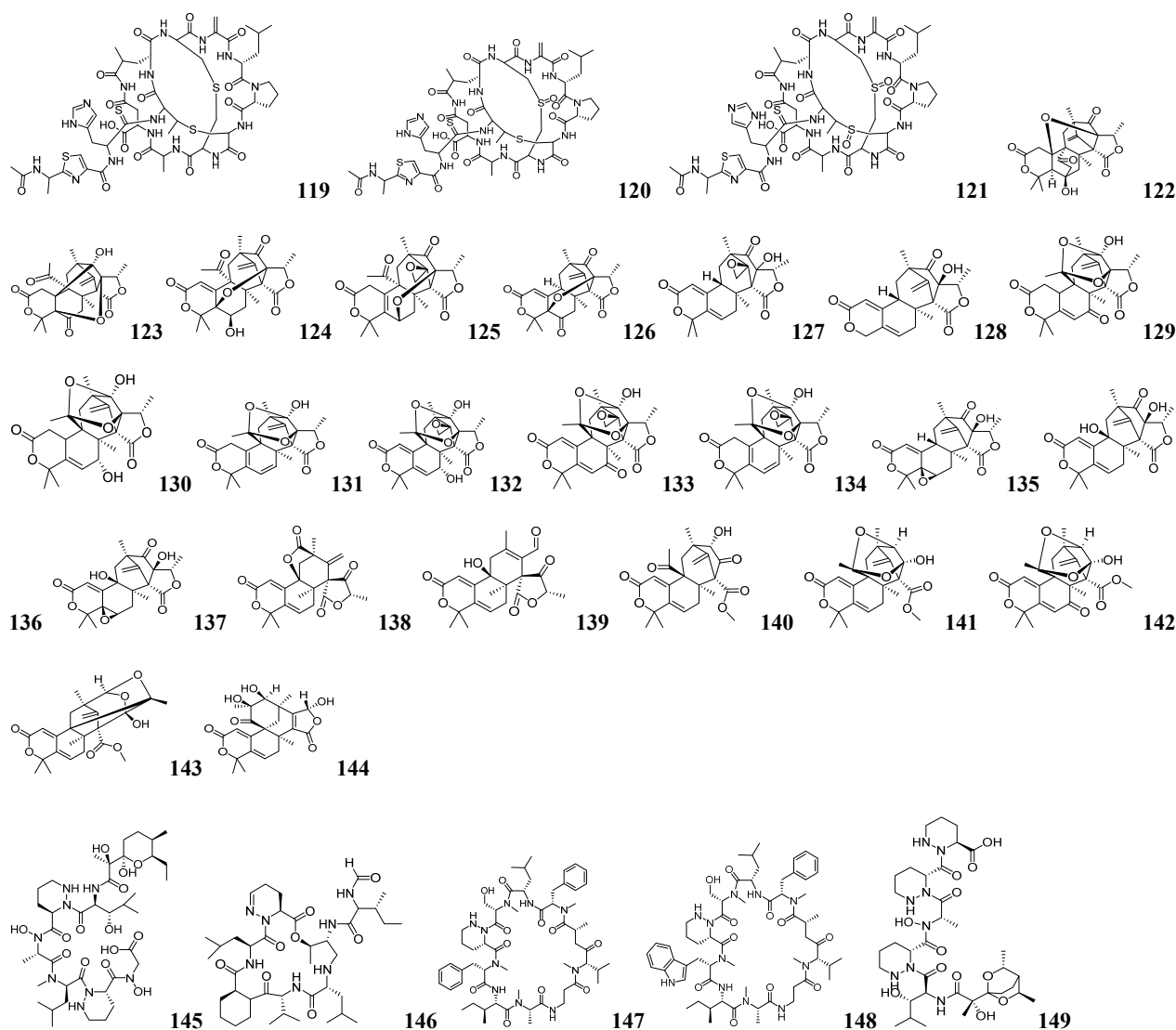


Figure 10. The chemical structures of NPs by integrated mining (119–149).

The SMART system (Secondary Metabolite Analysis Shell for Rapid Target identification) is a bioinformatics tool for microbial secondary metabolite target mining focused on the rapid identification and analysis of potential natural product BGCs [125]. It combines genome mining and metabolic network analysis strategies to accelerate the discovery and engineering of NPs. Ai-Li Fan and her team identified the *dhi* in the Antarctic fungus *Penicillium purpurogenum* AN13, which encodes dhilirane-type meroterpenoid (DM) analogs featuring a 3,5-dimethylorsellinic acid (DMOA) backbone [126]. A small-scale fermentation of *P. purpurogenum* AN13 was performed. Crude extracts were analyzed by NMR and LC-MS/MS, with HSQC data uploaded to the SMART system. Integration of GNPS and SMART analysis led to the identification of seven new DMs, namely dhilirolides O–U (122–128), as shown in Figure 10. SSN analysis of the key oxidase DhiD, which clusters with characterized α -KG-dependent oxygenases despite lacking close homology to known DMOA-pathway enzymes, demonstrated its exceptional catalytic versatility in generating structural diversity—a capacity confirmed through functional characterization that yielded 16 additional novel DMs (129–144) as shown in Figure 10. This research demonstrates how the SMART system's dual genome-metabolome analysis capability can accelerate NP discovery by efficiently connecting biosynthetic genes to their metabolic products.

Dong-Chan Oh and his team developed an effective strategy combining genomic and metabolic spectroscopic features to target the mining of piperazinic acid (Piz)-containing NPs [127]. For targeted genome mining, they designed degenerate primers (pizKtzI and pizKtzT) targeting conserved regions of the KtzT and KtzI biosynthetic genes. From 2020 screened strains, they identified 62 putative Piz-producing candidates. Phylogenetic analysis of KtzT/KtzI amplicons clustered these strains into distinct evolutionary clades, enabling branch-specific prediction of potential Piz-containing compound structures. Five representative strains—*Streptomyces* GB16, GSM11, PC5, SNJ018, and BYK1239—were selected as targets for further analysis. For metabolomic characterization, the screening strains were cultured with $^{15}\text{NH}_4\text{Cl}$ to isotopically label potential Piz compounds. Subsequent analysis by ^1H - ^{15}N -HSQC, HSQC-TOCSY, and HMBC confirmed the presence of Piz metabolites. Using the above strategy to isolate and characterize polyoxyperuvin B seco acid (**145**), depsidomycin D (**146**), and lenziamides A–B (**147**–**148**) as shown in Figure 10, Lenziamide A (**147**) has significant antiproliferative activity against colon cancer cells and overcomes 5-FU resistance. This study establishes a genomic and spectral characterization-based research approach that enables the targeted discovery of novel NPs containing specific structural motifs from microorganisms lacking complete genome sequences, overcoming the limitations of traditional methods.

Genome mining combined with NMR-guided strategies has wide applications in targeting the isolation of Piz-like compounds. Raymond J. Andersen and his team screened the NCBI database using the BLAST-P program with the amino acid sequence of the Piz synthase KtzT as a probe, ultimately selecting *Streptomyces incarnatus* NRRL 8089, a strain exhibiting over 50% homology [128]. The metabolic processes were labeled and analyzed by ^{15}N NMR, including ^1H - ^{15}N -HSQC and ^1H - ^{15}N -HSQC-TOCSY spectra of the crude extracts. The characteristic signals (δ_{N} 299.8 ppm and δ_{H} 4.52 ppm for N-H) and their associated spin systems in the Piz compounds guided a hierarchical separation, yielding the target incarnatapeptins A (**149**) and B (**150**) as shown in Figures 10 and 11. Similarly, Hua-Yue Li and his team analyzed the ^1H NMR data of the marine actinomycete *Streptomyces* sp. S063 and identified a characteristic negative hydrogen signal attributable to Piz [129]. Further analysis with antiSMASH 6.0 and SSN identified the Piz-related BGCs in this strain. Through NMR-guided stepwise isolation, two Piz-containing cyclic decapeptides, namely lenziamides D1 (**151**) and B1 (**152**), were obtained as shown in Figure 11. Combining genome mining with screening methods—including characteristic NMR signals and ^{15}N NMR—enables targeted isolation of low-abundance Piz-containing NPs, thereby expanding their structural diversity.

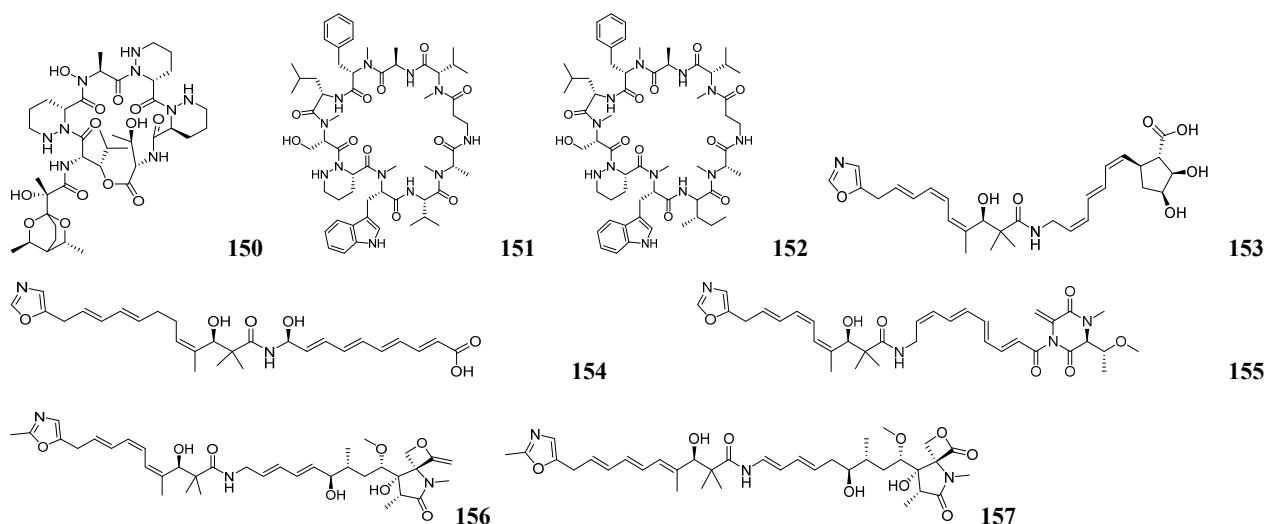


Figure 11. The chemical structures of NPs by integrated mining (**150**–**157**).

Dong-Chan Oh and his team have developed a new strategy that combines genomic characterization with spectral characterization to efficiently discover NPs containing specific structural motifs without the need for isotope labeling [130]. The specific mining process is shown in Figure 12A. During genomic characterization, PCR primers containing concatenated motifs (SGGKDS) were designed to screen for oxazole cyclase homologs, using known terminal oxazole compounds (inthomycin B and phthoxazolin) as references. In the genomic screening stage, degenerate PCR primers containing motifs (SGGKDS) were designed for screening the oxazole cyclase homologous genes based on the known terminal oxazole compounds inthomycin B and phthoxazolin. An in-house bacterial genomic DNA library of 1000 strains was screened and 16 hit strains were obtained and phylogenetically analyzed. To monitor oxazole production during small-scale fermentation, ^1H coupling was disabled during ^1H - ^{13}C HSQC acquisition, recording only $^1J_{\text{CH}}$ values and single-bond ^1H - ^{13}C correlations. The presence of oxazoles was confirmed by detecting characteristic single-bond ^1H - ^{13}C correlations at $\delta_{\text{C/H}}$ 151.1/8.24 ($^1J_{\text{CH}} = 230$ Hz) and $\delta_{\text{C/H}}$ 121.7/6.90 ($^1J_{\text{CH}} = 196$ Hz) in the ^1H - ^{13}C HSQC spectra. Five terminal oxazole compounds lenzioxazole (153), permafroxazole (154), tenebriazine (155), and methyl-oxazolomycins A–B (156–157) were isolated using ^1H - ^{13}C HSQC-guided fractionation as shown in Figure 11. This integrated approach combining genomic screening with ^1H - ^{13}C HSQC analysis enables efficient detection of terminal oxazole compounds without isotopic labeling. The method can be extended to identify other heterocyclic compounds, including discrimination between imidazoles and oxazoles, demonstrating broad applicability.

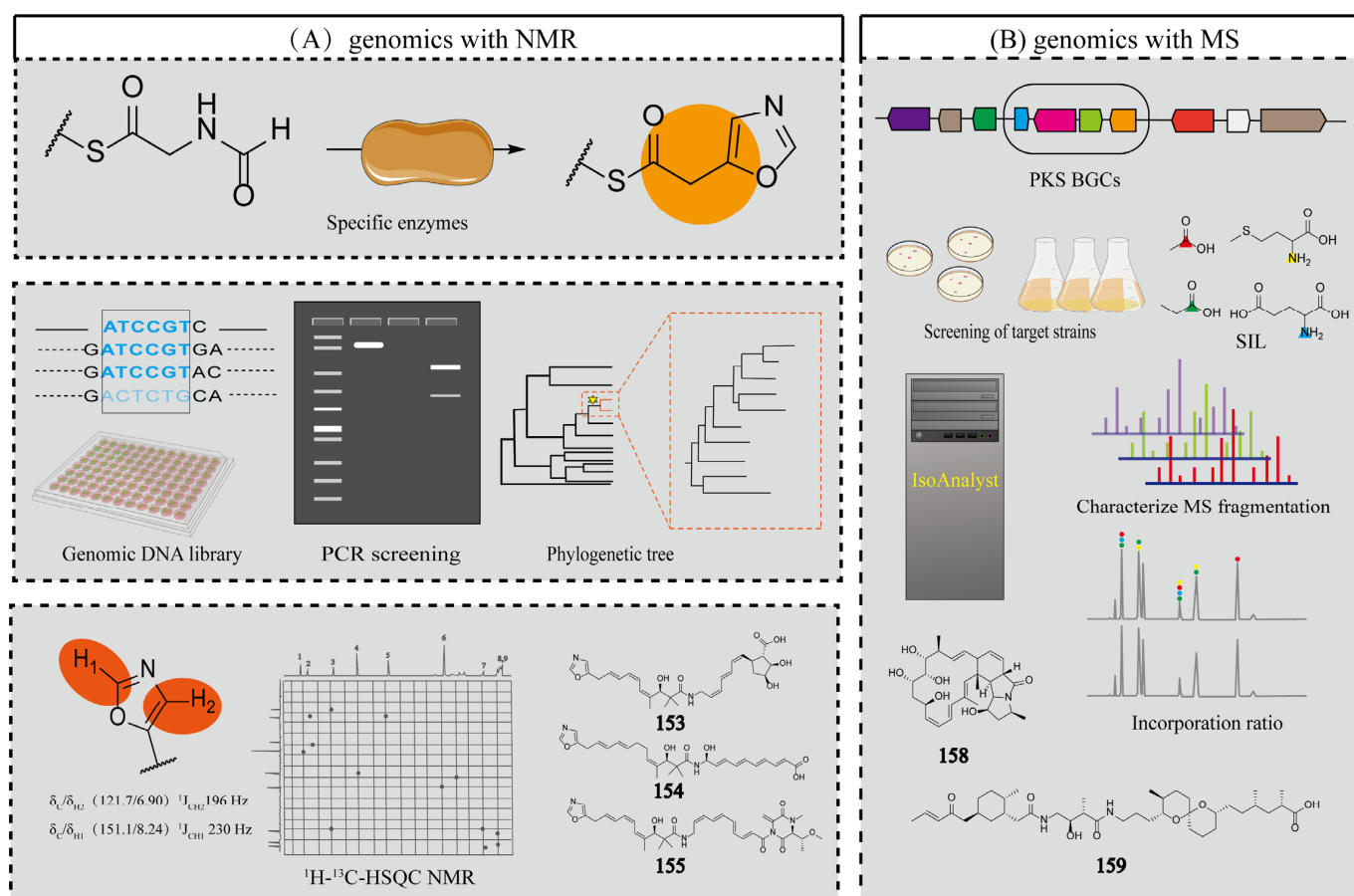


Figure 12. (A) Genomics and NMR aid in mining new NPs. (B) Genomics and MS aid in mining new NPs.

4.2. Genomics Combined with MS

MS is a powerful metabolomics tool offering high sensitivity, resolution, and rapid analysis. In 2014, Douglas A. Mitchell and his team successfully identified cyclo-thiazomycin C through an approach combining genomic screening with MS detection [131]. The core innovation of this research lies in employing genome mining to identify microbial strains capable of producing dehydroamino acids (DHAAs). Treatment of extracts with dithiothreitol (DTT) resulted in a 154.0 Da mass increase of DTT adducts compared to the output metabolite, a feature that can be targeted to track DHAAs analogs using MS monitoring. This study demonstrates that the integrated use of genomics and MS constitutes an efficient strategy for novel NPs discovery.

IsoAnalyst is an HR-MS-based software for isotopic fine structure (IFS) analysis. It can analyze isotope peak intensity ratios to determine the most probable molecular formula, while cross-referencing multiple databases enhances identification accuracy [132]. Roger G. Linington and his team utilized the IsoAnalyst platform to create metabolite-BGC mappings associated with BGC-based chemical structure predictions. The specific mining process is shown in Figure 12B. Specifically, four universal precursors—[1-¹³C] propionate, [1-¹³C] acetate, [1-¹⁵N] glutamate, and [methyl-¹³C] methionine—were used as stable isotope labeling (SIL) carriers to participate in biosynthetic processes as substrates. A control group without any additions was established to compare the MS data before and after treatment. The data were processed using IsoAnalyst to compare mass isotope distributions between labeled and unlabeled conditions, determining the degree of labeling for each precursor. Simultaneously, the strain genome was mined for BGCs using antiSMASH. Substrates for each BGC were predicted by integrating the MIBiG database with literature-derived information, generating a substrate prediction table. SIL precursor incorporation patterns were manually compared to theoretical incorporation rates predicted from BGC annotations to identify candidate BGCs associated with each labeled metabolite. The complete genome of *Micromonospora* sp. was analyzed using antiSMASH, and BGCs were manually curated for substrate prediction, generating a table of predicted BGC markers. The 246 MS features labeled under two or more SIL conditions were processed in IsoAnalyst, filtered to 100 based on peak type and intensity, and categorized into two broad groups according to their isotopic labeling patterns. Comparison with the annotated BGC list revealed that one of the most promising classes was associated with BGC 30c. BGC 30c was identified as responsible for the biosynthesis of lobosamide-like compounds. The predominant labeled metabolites in this group exhibited characteristic *m/z* profiles matching the lobosamide family, including distinctive [M+H–H₂O]⁺ fragment ions generated through intragenic cleavage. Based on these findings, subsequent isolation and purification yielded two known lobosamide analogs and a new natural product, which was designated lobosamide D (**158**) as shown in Figure 13.

Roger G. Linington and his team also used IsoAnalyst to discover the new polyketide compound lagriamide B (**159**) from a Burkholderiales strain as shown in Figure 13 [133]. The researchers first sequenced 115 *Burkholderia cepacia* strains and conducted bioinformatics analysis using antiSMASH 6.0, identifying a gene cluster homologous to the known antifungal polyketide lagriamide A BGC. They screened the NCBI database for additional strains containing this conserved gene cluster. Building on the structural features of lagriamide A, parallel stable isotope labeling experiments were conducted using [1-¹⁵N] glutamic acid as the SIL substrate in two strains harboring the target BGCs. IsoAnalyst was employed to preferentially screen for molecules containing the target elements or structural motifs. Lagriamide B (**159**) was successfully identified. The IsoAnalyst platform enables the classification of specific metabolites through SIL and facilitates the correlation of these compounds with candidate BGCs. IsoAnalyst prioritizes compound mining based

on biosynthetic relevance, employing a screening process that emphasizes biosynthetic relationships. Notably, it can identify biogenetic associations between molecules even when their spectral features differ significantly. This study demonstrates significant potential for application in the discovery of novel structural variants within compound families.

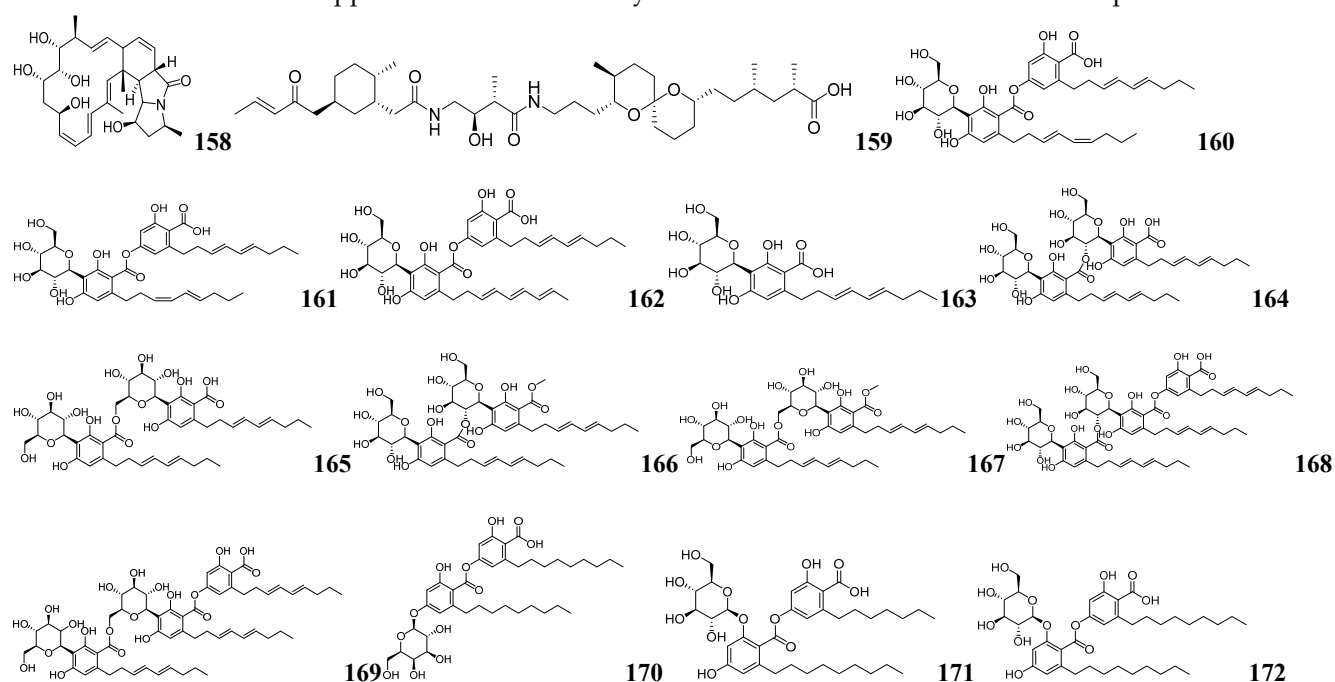


Figure 13. The chemical structures of NPs by integrated mining (158–172).

Jun-Gui Dai and his team developed an integrated approach combining enzyme-guided genome mining with MS-based metabolomics to discover multiple novel glycosylated NPs from fungal sources [134]. Functional characterization of stromemycin biosynthetic genes in *Aspergillus ustus* through in vivo gene knockout and heterologous expression revealed StmC as the responsible glycosyltransferase. Using AuCGT as a probe, they mined 158 homologous proteins and identified 80 candidate strains from NCBI and other databases. AntiSMASH analysis predicted that most AuCGT homologs co-localize with PKS clusters, suggesting their potential role in polyketide glycoside biosynthesis. To validate the polyketide glycoside synthesis potential of the candidate strains, seven fungal strains were preferred for fermentation culture, of which three strains were confirmed to possess the characteristic fragment ions of C-chain hexoses ($[M-H-120]^-$ and $[M-H-90]^-$). Through large-scale cultivation of *T. cellulolyticus*, 10 new glycosides, namely talarocellmycins A–C (160–162), carnemycin I (163), talarocellmycins D–G (164–167), and talarocellmycins H–I (168–169) were isolated. Eight novel glycosides, namely phaeomoniecin A (170), phaeomoniecin B–C (171–172), phaeomoniecins D–G (173–176) and phaeomoniecin H (177), were obtained from *P. chlamydospora*. Two new α -pyran-3-C- β -D-galactoside compounds, verapyrones A–B (178–179), were obtained from *V. dahliae*. The structures of these compounds are shown in Figures 13 and 14. This study also identified novel condensed phenolic acid C-/O-glycosyltransferases and α -pyranone C-glycosyltransferases, confirming that other fungal candidates, beyond the three mentioned strains, can also synthesize such glycosidic compounds. Genome mining with these novel glycosyltransferases could reveal more structurally distinct fungal glycosides.

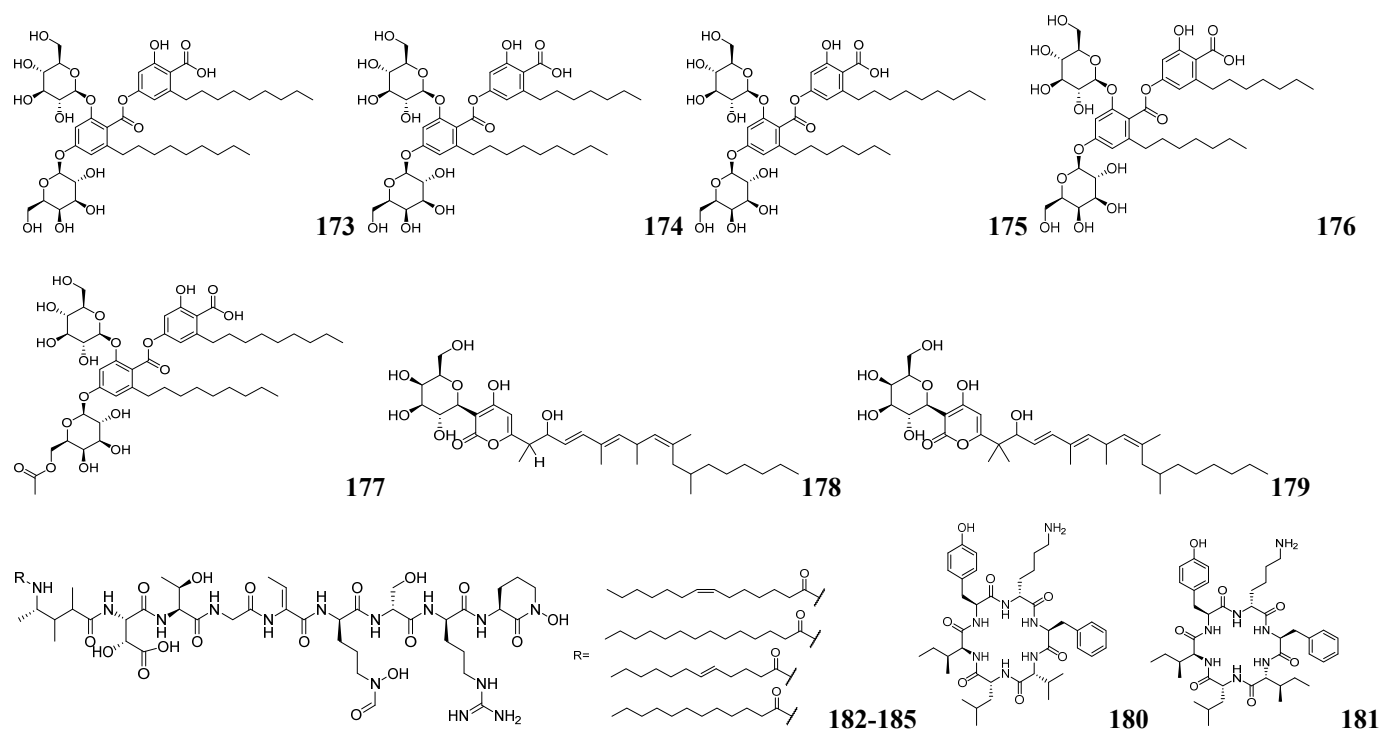


Figure 14. The chemical structures of NPs by integrated mining (173–185).

Integrated genome mining and MS/MS analysis are widely used approaches for discovering novel NPs. Seung Hyun Kim and his team isolated nine bacterial strains from the sea pony epiphytic environment [135]. Among them, YSL2 exhibited the strongest inhibitory activity and high phylogenetic novelty, as a new species of the genus *Nocardia*, named *Nocardiopsis maritima* sp. nov. Strain YSL2 harbors 17 BGCs, one of which is linked to hexapeptide synthesis and may produce novel NPs. The hexapeptide analog maritiamides A–B (180–181) were successfully traced and characterized by LC-MS/MS-based molecular network analysis as shown in Figure 14. Similarly, Paul D. Boudreau and his team identified delftibactins BGC-homologous gene clusters in *D. lacustris* through whole-genome sequencing and antiSMASH analysis [136]. Iron restriction induced *D. lacustris* to produce novel lipophilic metallophores. Through MS-guided isolation, they obtained four structurally distinct delftibactin derivatives, namely delftibactins C–F (182–185), as shown in Figure 14 featuring modified lipid tails. This study employs an integrated metabolomics-genomics approach to systematically investigate bacterial specialized metabolites, elucidating both their chemical structures and biosynthetic pathways.

5. Discussion

This review comprehensively summarizes 185 new natural products discovered with genomics, metabolomics, and their integration strategies from 2018 to 2024, while highlighting key technological breakthroughs and emerging research frontiers. In RiPPs mining, tools such as MSSN and SPECO were used to link P450 enzymes to RiPPs. Subsequently, AlphaFold-Multimer was used to predict the enzyme–substrate binding conformation and validate the catalytic mechanism of P450 enzymes for RiPPs. The availability of multiple search tools for specific BGCs has allowed researchers to conduct more detailed targeted analyses of a particular class of BGCs of interest. These efforts are undoubtedly more detailed and in-depth compared to pre-2018 tools. For terpene discovery, HMMER screened fungal genomes for PTTS, while FunBGCeX optimized gene cluster boundary prediction, addressing pre-2018 limitations in non-model organism gene annotation. In polyketide mining,

Seq2PKS predicted polyketide structures by integrating gene sequence–structure correlations, a strategy that expanded detectable BGC types compared to previous single-omics methods. This multi-dimensional validation (genomic prediction + metabolic profiling) minimized false negatives, enabling targeted isolation of low-abundance metabolites.

Metabolomics advancements, such as UPLC-Q-TOF-MS/MS, enabled high-resolution mass fragmentation for precise metabolite identification. The GNPS platform constructed molecular networks to cluster structurally related compounds, while FBMN (Fragmentation-Based Molecular Networking) incorporated retention time and peak intensity to resolve isomers—overcoming pre-2018 challenges in spectral ambiguity. SMAST (Structure Elucidation via Molecular Networking) directly linked NMR/HSQC data to BGCs, capturing mixture-specific features more efficiently than traditional metabolite tracing.

The integration of genomics and metabolomics strategies significantly improved NPs discovery efficiency through multidimensional technology synergies. We have summarized the genomics–metabolomics integration strategies mentioned in the text in a tabular format (as shown in Table 1) to show more clearly the workflow and their significant advantages. For example, in the discovery of nocathioamides A–C, RiPPs gene clusters containing thiazole/oxazole motifs were predicted for the first time from *Nocardia terpenica* genomes using antiSMASH and RODEO. Subsequently, stable isotope labeling was used to trace the peptide biosynthesis pathway, followed by identification and precise localization of the identified compounds by characteristic HMBC signals in complex metabolite mixtures. This integrated strategy not only overcomes the limitations of single-omics approaches for gene–metabolite associations, but also enables systematic metabolic pathway analysis through a closed-loop workflow combining bioinformatics predictions, spectroscopic validation, and isotope tracing. This approach accelerates the targeted discovery of low-abundance bioactive molecules and facilitates mechanistic elucidation. At the same time, several recent studies have shown that the integration strategy has facilitated the discovery of novel skeletons that have never been reported before, such as the discovery of the novel antifungal molecule mandimycin [54], the rare macrolide skeleton somalactams [137], and the non-squalene triterpene colleterpenol [138], to name just a few. The new discovery paradigm allows researchers to target the “dark matter” in microbial genomes, which will undoubtedly greatly accelerate the discovery of new skeletons and thus enrich the chemical diversity of natural molecules.

Although significant progress has been made for large-scale analysis of BGCs, direct prediction of chemical structures based on gene sequences is a difficult problem that is currently unattainable. The solution to this problem depends on further advances in synthetic biology, especially the accumulation of data for experimental validation of biosynthetic pathways. This will lead to an in-depth development of genomics analysis from BGCs type determination to compound structure prediction. MS-based metabolomics analyses also face the problem of false-positive detection signals and distorted prediction results. In addition to hardware performance upgrades, the performance of mass spectrometry fragment-based structural analysis may also be enhanced by AI-enabled big data modeling [139–142].

In summary, there is currently a profound paradigm shift in microbial natural products research, characterized by the emergence of synergistic, technology-driven discovery. While challenges in database integrity, strain culturability, and multi-omics integration remain, the combined power of these technological approaches has opened up unprecedented avenues for microbial chemical diversity. By refining predictive models, optimizing experimental workflows, and supporting interdisciplinary collaborations, the field is poised to overcome

current limitations and deliver a new generation of bioactive molecules, cementing the central role of microbial natural products in drug discovery for decades to come.

Table 1. The advantages of the integration of genomic and metabolomic strategies compared to before 2018.

Method	Workflow	Applications	Advancements
Genome mining combined with NMR-based isotope labeling	- antiSMASH, RODEO and RiPPMiner combined prediction of RiPPs-type BGCs. - Characteristic signals labeled by isotopes in ^1H-^{13}C HMBC spectra.	- RiPPs with characteristic functional groups such as thiazole, oxazole, imine, and thioamide [124]. - RiPPs with significant features in two-dimensional spectra are also applicable.	The comprehensive use of antiSMASH and various other tools enables more accurate identification of RiPPs BGC types.
Genome mining combined with SMART and molecular networking	- Establishing links between compounds and BGCs using SMART analysis of NMR and HSQC data. - Utilizing SSN to identify potential enzymes for subsequent expression.	DM class containing DMOA skeleton [125].	SMART has a stronger ability to capture characteristic information in mixtures and can directly associate compounds with BGC.
Genome mining combined with two-dimensional spectral features	- Design primers to screen homologous genes. - During the collection of ^1H-^{13}C HSQC, only the $^1J_{\text{CH}}$ value and the single bond correlation between ^1H and ^{13}C are recorded.	- Terminal oxazole NPs [130]. - NPs with special C-H correlations on ^1H-^{13}C HSQC.	Capturing single-key correlations enhances detection sensitivity, and a small trial is sufficient to determine whether a gene is successfully expressed.
Genome mining combined with isotope feeding and IsoAnalyst	- Utilize antiSMASH 6.0 predictions and compare them with MIBiG to identify potential BGCs. - Isotope-labeled precursor feeding. - IsoAnalyst analyzes its biosynthetic pathway and the approximate skeleton of the compound.	- Lobosamide-type polyene lactam [132]. - Lagriamide polyketides [133].	IsoAnalyst can accurately capture the connections between molecules that are biologically related but have significantly different spectral characteristics.

Author Contributions: Conceptualization, Z.W.; writing—original draft preparation, Z.W.; writing—review and editing, J.Y., Y.H. and C.W.; supervision, J.C. and H.W. All authors have read and agreed to the published version of the manuscript.

Funding: The project was supported by the National Key Research and Development Program of China (2022YFC2804205 and 2022YFC2804104), the Project of International Cooperation and Exchanges NSFC-ASRT (W2412100), the National Natural Science Foundation of China (42276137), and the Key Research and Development Program of Zhejiang Province (2021C03084). We also gratefully acknowledge platform support from the Zhejiang International SciTech Cooperation Base for the Exploitation and Utilization of Nature Product, Zhejiang Provincial Key Laboratory of TCM for Innovative R & D and Digital Intelligent Manufacturing of TCM Great Health Products, and Zhejiang Key Laboratory of Green, Low-Carbon and Efficient Development of Marine Fishery Resources.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Atanasov, A.G.; Zotchev, S.B.; Dirsch, V.M.; Supuran, C.T.; the International Natural Product Sciences Taskforce. Natural products in drug discovery: Advances and opportunities. *Nat. Rev. Drug Discov.* **2021**, *20*, 200–216. [\[CrossRef\]](#) [\[PubMed\]](#)
- Newman, D.J.; Cragg, G.M. Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* **2016**, *79*, 629–661. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ligon, B.L. Penicillin: Its discovery and early development. *Semin. Pediatr. Infect. Dis.* **2004**, *15*, 52–57. [\[CrossRef\]](#)
- Tribe, H.T. The discovery and development of cyclosporin. *Mycologist* **1998**, *12*, 20–22. [\[CrossRef\]](#)
- Tobert, J.A. Lovastatin and beyond: The history of the HMG-CoA reductase inhibitors. *Nat. Rev. Drug Discov.* **2003**, *2*, 517–526. [\[CrossRef\]](#)
- Chu, F.; Bai, Z.; Zhu, H. Research progress of microbial natural products against drug-resistant bacteria. *Nat. Prod. Res. Dev.* **2015**, *27*, 1466–1482.
- Milshcheyev, A.; Schneider, J.S.; Brady, S.F. Mining the metabiome: Identifying novel natural products from microbial communities. *Chem. Biol.* **2014**, *21*, 1211–1223. [\[CrossRef\]](#)
- Kalaitzis, J.A.; Ingre, S.D.; Chau, R.; Simon, Y.; Neilan, B.A. Genome-guided discovery of natural products and biosynthetic pathways from Australia's untapped microbial megadiversity. *Aust. J. Chem.* **2016**, *69*, 129–135. [\[CrossRef\]](#)
- Li, Z.; Zhu, D.; Shen, Y. Discovery of novel bioactive natural products driven by genome mining. *Drug Discov. Ther.* **2018**, *12*, 318–328. [\[CrossRef\]](#)
- Chi, H.; Liu, T. Synthetic biology promotes efficient production and innovative discovery of natural products. *Chin. Bull. Life Sci.* **2021**, *33*, 1510–1519.
- Sukmarini, L. Recent advances in discovery of lead structures from microbial natural products: Genomics- and metabolomics-guided acceleration. *Molecules* **2021**, *26*, 2542. [\[CrossRef\]](#)
- Wenger, A.M.; Peluso, P.; Rowell, W.J.; Chang, P.C.; Hall, R.J.; Concepcion, G.T.; Ebler, J.; Fungtammasan, A.; Kolesnikov, A.; Olson, N.D.; et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* **2019**, *37*, 1155–1162. [\[CrossRef\]](#) [\[PubMed\]](#)
- Loman, N.; Goodwin, S.; Jansen, H.; Loose, M. A disruptive sequencer meets disruptive publishing. *F1000Research* **2015**, *4*, 1074. [\[CrossRef\]](#) [\[PubMed\]](#)
- Rigali, S.; Anderssen, S.; Naomé, A.; van Wezel, G.P. Cracking the regulatory code of biosynthetic gene clusters as a strategy for natural product discovery. *Biochem. Pharmacol.* **2018**, *153*, 24–34. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kandiah, M.; Urban, P.L. Advances in ultrasensitive mass spectrometry of organic molecules. *Chem. Soc. Rev.* **2013**, *42*, 5299–5322. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kovacs, H.; Moskau, D.; Spraul, M. Cryogenically cooled probes—A leap in NMR technology. *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *46*, 131–155. [\[CrossRef\]](#)
- Habib, F.; Tocher, D.A.; Carmalt, C.J. Applications of the crystalline sponge method and developments of alternative crystalline sponges. *Mater. Today Proc.* **2022**, *56*, 3766–3773. [\[CrossRef\]](#)
- Li, J.; Liu, J.K.; Wang, W.X. GIAO ¹³C NMR calculation with sorted training sets improves accuracy and reliability for structural assignment. *J. Org. Chem.* **2020**, *85*, 11350–11358. [\[CrossRef\]](#) [\[PubMed\]](#)
- Pescitelli, G.; Di Bari, L.; Bero, N. Conformational aspects in the studies of organic compounds by electronic circular dichroism. *Chem. Soc. Rev.* **2011**, *40*, 4603–4625. [\[CrossRef\]](#)
- Sanger, F.; Air, G.M.; Barrell, B.G.; Brown, N.L.; Coulson, A.R.; Fiddes, C.A.; Hutchison, C.A.; Slocumbe, P.M.; Smith, M. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* **1977**, *265*, 687–695. [\[CrossRef\]](#)
- Auslander, N.; Gusow, A.B.; Koonin, E.V. Incorporating machine learning into established bioinformatics frameworks. *Int. J. Mol. Sci.* **2021**, *22*, 2903. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wei, S.; Wang, S. Structural stability-aware deep learning: Advancing RNA secondary structure prediction. In Proceedings of the 2024 Fourth International Conference on Biomedicine and Bioinformatics Engineering, Kaifeng, China, 14–16 June 2024. 132521G.
- Liu, Z.; Su, L.; Fang, X.; Chang, D.; Chen, Z.; Jiang, X.; Li, T.; Wang, Y.; Guo, Y.; Wang, J. A Spatial Strain of *Staphylococcus aureus* LCT-SA67. CN103087943B, 4 March 2015.
- Guo, Y.; Chang, D.; Chen, Z.; Wang, Y.; Su, L.; Wang, L.; Wang, J.; Liu, Z.; Li, T.; Fang, X. A Spatial Strain of *Klebsiella pneumoniae* LCT-KP289. CN102994414B, 12 November 2014.
- Bao, Z.; Hu, J.; Zhang, L.; Bao, L.; Yu, H.; Li, Y.; Wang, S. Integrating Micro- and Macro-scale Comparative Genomics Analysis Methods. CN117976041A, 3 May 2024.
- Blin, K.; Shaw, S.; Augustijn, H.E.; Reitz, Z.L.; Biermann, F.; Alanjary, M.; Fetter, A.; Terlouw, B.R.; Metcalf, W.W.; Helfrich, E.J.N.; et al. antiSMASH 7.0: New and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.* **2023**, *51*, W46–W50. [\[CrossRef\]](#)
- Hannigan, G.D.; Prihoda, D.; Palicka, A.; Soukup, J.; Klempir, O.; Rampula, L.; Durcak, J.; Wurst, M.; Kotowski, J.; Chang, D.; et al. A deep learning genome-mining strategy for biosynthetic gene cluster prediction. *Nucleic Acids Res.* **2019**, *47*, e110. [\[CrossRef\]](#)

28. Liu, M.Y.; Li, Y.; Li, H.Z. Deep learning to predict the biosynthetic gene clusters in bacterial genomes. *J. Mol. Biol.* **2022**, *434*, 167597. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Siامي-Namini, S.; Tavakoli, N.; Namin, A.S. The performance of LSTM and BiLSTM in forecasting time series. In Proceedings of the IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 3285–3292.
30. Luo, L.; Yang, Z.H.; Yang, P.; Zhang, Y.; Wang, L.; Lin, H.F.; Wang, J. An attention-based BiLSTM-CRF approach to document-level chemical named entity recognition. *Bioinformatics* **2018**, *34*, 1381–1388. [\[CrossRef\]](#)
31. Ren, Q.; Cheng, H.; Han, H. Research on machine learning framework based on random forest algorithm. In Proceedings of the International Conference on Advances in Materials, Machinery, Electronics (AMME), Wuhan, China, 25–26 February 2017.
32. Schonlau, M.; Zou, R.Y. The random forest algorithm for statistical learning. *Stata J.* **2020**, *20*, 3–29. [\[CrossRef\]](#)
33. Anker, A.S.; Friis-Jensen, U.; Johansen, F.L.; Billinge, S.J.L.; Jensen, K.M.Ø. *ClusterFinder*: A fast tool to find cluster structures from pair distribution function data. *Acta Crystallogr. A-Found. Adv.* **2024**, *80*, 213–220. [\[CrossRef\]](#)
34. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernysky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [\[CrossRef\]](#) [\[PubMed\]](#)
35. van Dijk, E.L.; Auger, H.; Jaszczyszyn, Y.; Thermes, C. Ten years of next-generation sequencing technology. *Trends Genet.* **2014**, *30*, 418–426. [\[CrossRef\]](#)
36. Heidersbach, A.J.; Dorigi, K.M.; Gomez, J.A.; Jacobi, A.M.; Haley, B. A versatile, high-efficiency platform for CRISPR-based gene activation. *Nat. Commun.* **2023**, *14*, 902. [\[CrossRef\]](#) [\[PubMed\]](#)
37. DeCarlo, P.F.; Kimmel, J.R.; Trimborn, A.; Northway, M.J.; Jayne, J.T.; Aiken, A.C.; Gonin, M.; Fuhrer, K.; Horvath, T.; Docherty, K.S.; et al. Field-deployable, high-resolution, time-of-flight aerosol mass spectrometer. *Anal. Chem.* **2006**, *78*, 8281–8289. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Marshall, A.G.; Hendrickson, C.L.; Jackson, G.S. Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom. Rev.* **1998**, *17*, 1–35. [\[CrossRef\]](#)
39. Olsen, J.V.; de Godoy, L.M.F.; Li, G.Q.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteom.* **2005**, *4*, 2010–2021. [\[CrossRef\]](#)
40. Comisarow, M.B.; Marshall, A.G. Fourier transform ion cyclotron resonance spectroscopy. *Chem. Phys. Lett.* **1974**, *25*, 282–283. [\[CrossRef\]](#)
41. Dumez, J.N. NMR methods for the analysis of mixtures. *Chem. Commun.* **2022**, *58*, 13855–13872. [\[CrossRef\]](#)
42. Shapiro, M.J.; Wareing, J.R. NMR methods in combinatorial chemistry. *Curr. Opin. Chem. Biol.* **1998**, *2*, 372–375. [\[CrossRef\]](#)
43. Lhoste, C.; Lorandel, B.; Praud, C.; Marchand, A.; Mishra, R.; Dey, A.; Bernard, A.; Dumez, J.N.; Giraudeau, P. Ultrafast 2D NMR for the analysis of complex mixtures. *Prog. Nucl. Magn. Reson. Spectrosc.* **2022**, *130*, 1–46. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Hansen, A.L.; Kupce, E.; Li, D.W.; Bruschweiler-Li, L.; Wang, C.; Brüschweiler, R. 2D NMR-based metabolomics with HSQC/TOCSY NOAH supersequences. *Anal. Chem.* **2021**, *93*, 6112–6119. [\[CrossRef\]](#)
45. Wang, M.X.; Carver, J.J.; Phelan, V.V.; Sanchez, L.M.; Garg, N.; Peng, Y.; Nguyen, D.D.; Watrous, J.; Kapon, 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. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Nothias, L.F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; Protsyuk, I.; Ernst, M.; Tsugawa, H.; Fleischauer, M.; et al. Feature-based molecular networking in the GNPS analysis environment. *Nat. Methods* **2020**, *17*, 905–908. [\[CrossRef\]](#)
47. Dührkop, K.; Fleischauer, M.; Ludwig, M.; Aksenov, A.A.; Melnik, A.V.; Meusel, M.; Dorrestein, P.C.; Rousu, J.; Bocker, S. SIRIUS 4: A rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Methods* **2019**, *16*, 299–302. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Deng, Y.J.; Yao, Y.; Wang, Y.N.; Yu, T.T.; Cai, W.H.; Zhou, D.L.; Yin, F.; Liu, W.L.; Liu, Y.Y.; Xie, C.B.; et al. An end-to-end deep learning method for mass spectrometry data analysis to reveal disease-specific metabolic profiles. *Nat. Commun.* **2024**, *15*, 7136. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Eckert-Boulet, N.; Nielsen, P.S.; Friis, C.; dos Santos, M.M.; Nielsen, J.; Kielland-Brandt, M.C.; Regenber, B. Transcriptional profiling of extracellular amino acid sensing in *Saccharomyces cerevisiae* and the role of Stp1p and Stp2p. *Yeast* **2004**, *21*, 635–648. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Ishii, N.; Nakahigashi, K.; Baba, T.; Robert, M.; Soga, T.; Kanai, A.; Hirasawa, T.; Naba, M.; Hirai, K.; Hoque, A.; et al. Multiple high-throughput analyses monitor the response of *E.coli* to perturbations. *Science* **2007**, *316*, 593–597. [\[CrossRef\]](#)
51. Huo, L.J.; Hug, J.J.; Fu, C.Z.; Bian, X.Y.; Zhang, Y.M.; Müller, R. Heterologous expression of bacterial natural product biosynthetic pathways. *Nat. Prod. Rep.* **2019**, *36*, 1412–1436. [\[CrossRef\]](#)
52. Romano, S.; Jackson, S.A.; Patry, S.; Dobson, A.D.W. Extending the “one strain many compounds” (OSMAC) principle to marine microorganisms. *Mar. Drugs* **2018**, *16*, 244. [\[CrossRef\]](#)

53. Si, Y.; Feng, Q. Application of OSMAC strategy in the study of microbial secondary metabolites. *J. Shenyang Pharm. Univ.* **2023**, *40*, 370–380.
54. Deng, Q.S.; Li, Y.C.; He, W.Y.; Chen, T.; Liu, N.; Ma, L.M.; Qiu, Z.X.; Shang, Z.; Wang, Z.Q. A polyene macrolide targeting phospholipids in the fungal cell membrane. *Nature* **2025**, *640*, 743–751. [\[CrossRef\]](#)
55. Montalbán-López, M.; Scott, T.A.; Ramesh, S.; Rahman, I.R.; van Heel, A.J.; Viel, J.H.; Bandarian, V.; Dittmann, E.; Genilloud, O.; Goto, Y.; et al. New developments in RiPP discovery, enzymology and engineering. *Nat. Prod. Rep.* **2021**, *38*, 130–239. [\[CrossRef\]](#)
56. Liu, J.; Liu, R.; He, B.-B.; Lin, X.; Guo, L.; Wu, G.; Li, Y.-X. Bacterial cytochrome P450 catalyzed macrocyclization of ribosomal peptides. *ACS Bio Med Chem Au* **2024**, *4*, 268–279. [\[CrossRef\]](#)
57. Kunakom, S.; Otani, H.; Uduary, D.W.; Doering, D.T.; Mouncey, N.J. Cytochromes P450 involved in bacterial RiPP biosyntheses. *J. Ind. Microbiol. Biotechnol.* **2023**, *50*, 2023. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Zhong, G. Cytochromes P450 associated with the biosyntheses of ribosomally synthesized and post-translationally modified peptides. *ACS Bio Med Chem Au* **2023**, *3*, 371–388. [\[CrossRef\]](#)
59. Laws, D., III; Plouch, E.V.; Blakey, S.B. Synthesis of ribosomally synthesized and post-translationally modified peptides containing C-C cross-links. *J. Nat. Prod.* **2022**, *85*, 2519–2539. [\[CrossRef\]](#) [\[PubMed\]](#)
60. Zhu, W.S.; Shenoy, A.; Kundrotas, P.; Elofsson, A. Evaluation of alphafold-multimer prediction on multi-chain protein complexes. *Bioinformatics* **2023**, *39*, btad424. [\[CrossRef\]](#)
61. He, B.B.; Liu, J.; Cheng, Z.; Liu, R.Z.; Zhong, Z.; Gao, Y.; Liu, H.Y.; Song, Z.M.; Tian, Y.Q.; Li, Y.X. Bacterial cytochrome P450 catalyzed post-translational macrocyclization of ribosomal peptides. *Angew. Chem.-Int. Ed.* **2023**, *62*, e202311533. [\[CrossRef\]](#)
62. Gerlt, J.A.; Bouvier, J.T.; Davidson, D.B.; Imker, H.J.; Sadkhin, B.; Slater, D.R.; Whalen, K.L. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta-Proteins Proteom.* **2015**, *1854*, 1019–1037. [\[CrossRef\]](#)
63. Moffat, A.D.; Santos-Aberturas, J.; Chandra, G.; Truman, A.W. A user guide for the identification of new RiPP biosynthetic gene clusters using a RiPPER-based workflow. In *Antimicrobial Therapies: Methods and Protocols*; Barreiro, C., Barredo, J.L., Eds.; Springer: Berlin/Heidelberg, Germany, 2021; Volume 2296, pp. 227–247.
64. Santos-Aberturas, J.; Chandra, G.; Frattaruolo, L.; Lacret, R.; Pham, T.H.; Vior, N.M.; Eyles, T.H.; Truman, A.W. Uncovering the unexplored diversity of thioamidated ribosomal peptides in *Actinobacteria* using the RiPPER genome mining tool. *Nucleic Acids Res.* **2019**, *47*, 4624–4637. [\[CrossRef\]](#) [\[PubMed\]](#)
65. Liu, C.L.; Wang, Z.J.; Shi, J.; Yan, Z.Y.; Zhang, G.D.; Jiao, R.H.; Tan, R.X.; Ge, H.M. P450-modified multicyclic cyclophane-containing ribosomally synthesized and post-translationally modified peptides. *Angew. Chem.-Int. Ed.* **2024**, *63*, e202314046. [\[CrossRef\]](#)
66. Altschul, S.; Madden, T.; Schaffer, A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *FASEB J.* **1998**, *12*, A1326. [\[CrossRef\]](#) [\[PubMed\]](#)
67. 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. [\[CrossRef\]](#)
68. Nam, H.; An, J.S.; Lee, J.; Yun, Y.; Lee, H.; Park, H.; Jung, Y.; Oh, K.B.; Oh, D.C.; Kim, S. Exploring the diverse landscape of biaryl-containing peptides generated by cytochrome P450 macrocyclases. *J. Am. Chem. Soc.* **2023**, *145*, 22047–22057. [\[CrossRef\]](#) [\[PubMed\]](#)
69. Wenzel, S.C.; Müller, R. Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways. *Curr. Opin. Biotechnol.* **2005**, *16*, 594–606. [\[CrossRef\]](#) [\[PubMed\]](#)
70. Gomez-Escribano, J.P.; Bibb, M.J. Heterologous expression of natural product biosynthetic gene clusters in *Streptomyces coelicolor*: From genome mining to manipulation of biosynthetic pathways. *J. Ind. Microbiol. Biotechnol.* **2014**, *41*, 425–431. [\[CrossRef\]](#)
71. Caldwell, B.J.; Bell, C.E. Structure and mechanism of the red recombination system of bacteriophage λ . *Prog. Biophys. Mol. Biol.* **2019**, *147*, 33–46. [\[CrossRef\]](#)
72. Murphy, K.C. λ recombination and recombineering. *EcoSal Plus* **2016**, *7*. [\[CrossRef\]](#) [\[PubMed\]](#)
73. Kawahara, T.; Izumikawa, M.; Kozono, I.; Hashimoto, J.; Kagaya, N.; Koiwai, H.; Komatsu, M.; Fujie, M.; Sato, N.; Ikeda, H.; et al. Neothioviridamide, a polythioamide compound produced by heterologous expression of a *Streptomyces* sp. Cryptic RiPP biosynthetic gene cluster. *J. Nat. Prod.* **2018**, *81*, 264–269. [\[CrossRef\]](#)
74. Guo, M.X.; Zhang, M.M.; Sun, K.; Cui, J.J.; Liu, Y.C.; Gao, K.; Dong, S.H.; Luo, S.W. Genome mining of linaridins provides insights into the widely distributed LinC oxidoreductases. *J. Nat. Prod.* **2023**, *86*, 2333–2341. [\[CrossRef\]](#)
75. Parson, W.; Strobl, C.; Huber, G.; Zimmermann, B.; Gomes, S.M.; Souto, L.; Fendt, L.; Delpont, R.; Langit, R.; Wootton, S.; et al. Evaluation of next generation mtGenome sequencing using the Ion Torrent Personal Genome Machine (PGM). *Forensic Sci. Int.-Genet.* **2013**, *7*, 543–549. [\[CrossRef\]](#)
76. Aziz, R.K.; Bartels, D.; Best, A.A.; DeJongh, M.; Disz, T.; Edwards, R.A.; Formsma, K.; Gerdes, S.; Glass, E.M.; Kubal, M.; et al. The RAST server: Rapid annotations using subsystems technology. *BMC Genom.* **2008**, *9*, 75. [\[CrossRef\]](#)

77. Cortés-Albayay, C.; Jarmusch, S.A.; Trusch, F.; Ebel, R.; Andrews, B.A.; Jaspars, M.; Asenjo, J.A. Downsizing class II lasso peptides: Genome mining-guided isolation of huascopeptin containing the first gly1-asp7 macrocycle. *J. Org. Chem.* **2020**, *85*, 1661–1667. [\[CrossRef\]](#)
78. Lei, R.; Tao, H.; Liu, T. Deep genome mining boosts the discovery of microbial terpenoids. *Synth. Biol. J.* **2024**, *5*, 507–526.
79. Chen, R.; Jia, Q.D.; Mu, X.; Hu, B.; Sun, X.; Deng, Z.X.; Chen, F.; Bian, G.K.; Liu, T.G. Systematic mining of fungal chimeric terpene synthases using an efficient precursor-providing yeast chassis. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2023247118. [\[CrossRef\]](#) [\[PubMed\]](#)
80. He, H.B.; Bian, G.K.; Herbst-Gervasoni, C.J.; Mori, T.; Shinsky, S.A.; Hou, A.W.; Mu, X.; Huang, M.J.; Cheng, S.; Deng, Z.X.; et al. Discovery of the cryptic function of terpene cyclases as aromatic prenyltransferases. *Nat. Commun.* **2020**, *11*, 3958. [\[CrossRef\]](#)
81. Chen, C.C.; Malwal, S.R.; Han, X.; Liu, W.D.; Ma, L.X.; Zhai, C.; Dai, L.H.; Huang, J.W.; Shillo, A.; Desai, J.; et al. Terpene cyclases and prenyltransferases: Structures and mechanisms of action. *ACS Catal.* **2021**, *11*, 290–303. [\[CrossRef\]](#)
82. Yu, D.S.; Lee, D.H.; Kim, S.K.; Lee, C.H.; Song, J.Y.; Kong, E.B.; Kim, J.F. Algorithm for predicting functionally equivalent proteins from BLAST and HMMER searches. *J. Microbiol. Biotechnol.* **2012**, *22*, 1054–1058. [\[CrossRef\]](#) [\[PubMed\]](#)
83. Hubley, R.; Finn, R.D.; Clements, J.; Eddy, S.R.; Jones, T.A.; Bao, W.D.; Smit, A.F.A.; Wheelers, T.J. The Dfam database of repetitive DNA families. *Nucleic Acids Res.* **2016**, *44*, D81–D89. [\[CrossRef\]](#) [\[PubMed\]](#)
84. Sayers, E.W.; Beck, J.; Bolton, E.E.; Brister, J.R.; Chan, J.; Comeau, D.C.; Connor, R.; DiCuccio, M.; Farrell, C.M.; Feldgarden, M.; et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **2024**, *52*, D33–D43. [\[CrossRef\]](#)
85. Brown, G.R.; Hem, V.; Katz, K.S.; Ovetsky, M.; Wallin, C.; Ermolaeva, O.; Tolstoy, I.; Tatusova, T.; Pruitt, K.D.; Maglott, D.R.; et al. Gene: A gene-centered information resource at NCBI. *Nucleic Acids Res.* **2015**, *43*, D36–D42. [\[CrossRef\]](#)
86. Bateman, A.; Martin, M.J.; O'Donovan, C.; Magrane, M.; Apweiler, R.; Alpi, E.; Antunes, R.; Ar-Ganiska, J.; Bely, B.; Bingley, M.; et al. UniProt: A hub for protein information. *Nucleic Acids Res.* **2015**, *43*, D204–D212.
87. Bateman, A.; Martin, M.J.; Orchard, S.; Magrane, M.; Adesina, A.; Ahmad, S.; Bowler-Barnett, E.H.; Bye-A-Jee, H.; Carpentier, D.; Denny, P.; et al. UniProt: The Universal Protein Knowledgebase in 2025. *Nucleic Acids Res.* **2024**, *52*, D609–D617.
88. Tang, J.; Matsuda, Y. Discovery of fungal onoceroide triterpenoids through domainless enzyme-targeted global genome mining. *Nat. Commun.* **2024**, *15*, 4312. [\[CrossRef\]](#) [\[PubMed\]](#)
89. Racolta, S.; Juhl, P.B.; Sirim, D.; Pleiss, J. The triterpene cyclase protein family: A systematic analysis. *Proteins-Struct. Funct. Bioinform.* **2012**, *80*, 2009–2019. [\[CrossRef\]](#)
90. Chen, R.; Feng, T.; Li, M.; Zhang, X.Y.; He, J.; Hu, B.; Deng, Z.X.; Liu, T.A.; Liu, J.K.; Wang, X.H.; et al. Characterization of tremulane sesquiterpene synthase from the basidiomycete *Irpex lacteus*. *Org. Lett.* **2022**, *24*, 5669–5673. [\[CrossRef\]](#) [\[PubMed\]](#)
91. Li, Z.; Jiang, Y.Y.; Zhang, X.W.; Chang, Y.M.; Li, S.; Zhang, X.M.; Zheng, S.M.; Geng, C.; Men, P.; Ma, L.; et al. Fragrant venezuelaenes A and B with A 5-5-6-7 tetracyclic skeleton: Discovery, biosynthesis, and mechanisms of central catalysts. *ACS Catal.* **2020**, *10*, 5846–5851. [\[CrossRef\]](#)
92. Zhang, P.; Wu, G.W.; Heard, S.C.; Niu, C.S.; Bell, S.A.; Li, F.L.; Ye, Y.; Zhang, Y.H.; Winter, J.M. Identification and characterization of a cryptic bifunctional type I diterpene synthase involved in talaronoid biosynthesis from a marine-derived fungus. *Org. Lett.* **2022**, *24*, 7037–7041. [\[CrossRef\]](#) [\[PubMed\]](#)
93. Sun, X.; Cai, Y.S.; Yuan, Y.J.; Bian, G.K.; Ye, Z.L.; Deng, Z.X.; Liu, T.G. Genome mining in *Trichoderma viride* J1-030: Discovery and identification of novel sesquiterpene synthase and its products. *Beilstein J. Org. Chem.* **2019**, *15*, 2052–2058. [\[CrossRef\]](#)
94. Li, W.Z.; Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [\[CrossRef\]](#)
95. Fu, L.M.; Niu, B.F.; Zhu, Z.W.; Wu, S.T.; Li, W.Z. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **2012**, *28*, 3150–3152. [\[CrossRef\]](#)
96. Navarro-Muñoz, J.C.; Selem-Mojica, N.; Mullowney, M.W.; Kautsar, S.A.; Tryon, J.H.; Parkinson, E.; 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. [\[CrossRef\]](#)
97. Liu, W.C.; Tian, X.Y.; Huang, X.; Malit, J.J.L.; Wu, C.H.; Guo, Z.H.; Tang, J.W.; Qian, P.Y. Discovery of P450-modified sesquiterpenoids levinoids A-D through global genome mining. *J. Nat. Prod.* **2024**, *87*, 876–883. [\[CrossRef\]](#)
98. Guo, J.J.; Cai, Y.S.; Cheng, F.C.; Yang, C.J.; Zhang, W.Q.; Yu, W.L.; Yan, J.J.; Deng, Z.X.; Hong, K. Genome mining reveals a multiproduct sesterterpenoid biosynthetic gene cluster in *Aspergillus ustus*. *Org. Lett.* **2021**, *23*, 1525–1529. [\[CrossRef\]](#) [\[PubMed\]](#)
99. Yan, D.H.; Zhou, M.Q.; Adduri, A.; Zhuang, Y.H.; Guler, M.; Liu, S.T.; Shin, H.; Kovach, T.; Oh, G.; Liu, X.; et al. Discovering type I cis-AT polyketides through computational mass spectrometry and genome mining with Seq2PKS. *Nat. Commun.* **2024**, *15*, 5356. [\[CrossRef\]](#) [\[PubMed\]](#)
100. Liaw, Y.C. Improvement of the fast exact pairwise-nearest-neighbor algorithm. *Pattern Recognit.* **2009**, *42*, 867–870. [\[CrossRef\]](#)
101. 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. [\[CrossRef\]](#)

102. Zhao, L.Y.; Shi, J.; Xu, Z.Y.; Sun, J.L.; Yan, Z.Y.; Tong, Z.W.; Tan, R.X.; Jiao, R.H.; Ge, H.M. Hybrid type I and II polyketide synthases yield distinct aromatic polyketides. *J. Am. Chem. Soc.* **2024**, *146*, 29462–29468. [\[CrossRef\]](#) [\[PubMed\]](#)
103. Dong, J.Y.; Tang, M.C.; Liu, L. α -pyrone derivatives from *Calcarisporium arbuscula* discovered by genome mining. *J. Nat. Prod.* **2023**, *86*, 2496–2501. [\[CrossRef\]](#)
104. Khan, H.; Ali, J. UHPLC/Q-TOF-MS Technique: Introduction and Applications. *Lett. Org. Chem.* **2015**, *12*, 371–378. [\[CrossRef\]](#)
105. Alsaleh, M.; Barbera, T.A.; Andrews, R.H.; Sithithaworn, P.; Khuntikeo, N.; Loilome, W.; Yongvanit, P.; Cox, I.J.; Syms, R.R.A.; Holmes, E.; et al. Mass spectrometry: A guide for the clinician. *J. Clin. Exp. Hepatol.* **2019**, *9*, 597–606. [\[CrossRef\]](#)
106. Liu, J.Z.; Wang, Y.D.; Fang, H.Q.; Sun, G.B.; Ding, G. UPLC-Q-TOF-MS/MS-based targeted discovery of chetomin analogues from *Chaetomium cochliodes*. *J. Nat. Prod.* **2024**, *87*, 1660–1665. [\[CrossRef\]](#)
107. Hu, Y.W.; Ma, S.; Pang, X.Y.; Cong, M.J.; Liu, Q.Q.; Han, F.H.; Wang, J.J.; Feng, W.E.; Liu, Y.H.; Wang, J.F. Cytotoxic pyridine alkaloids from a marine-derived fungus *Arthrimum arundinis* exhibiting apoptosis-inducing activities against small cell lung cancer. *Phytochemistry* **2023**, *213*, 113765. [\[CrossRef\]](#)
108. Chen, C.M.; Chen, W.H.; Tao, H.M.; Yang, B.; Zhou, X.F.; Luo, X.W.; Liu, Y.H. Diversified polyketides and nitrogenous compounds from the mangrove endophytic fungus *Penicillium steckii* SCSIO 41025. *Chin. J. Chem.* **2021**, *39*, 2132–2140. [\[CrossRef\]](#)
109. Guo, H.J.; Daniel, J.M.; Seibel, E.; Burkhardt, I.; Conlon, B.H.; Görls, H.; Vassao, D.G.; Dickschat, J.S.; Poulsen, M.; Beemelmans, C. Insights into the metabolomic capacity of *Podaxis* and isolation of podaxisterols A-D, ergosterol derivatives carrying nitrosyl cyanide-derived modifications. *J. Nat. Prod.* **2022**, *85*, 2159–2167. [\[CrossRef\]](#)
110. Marner, M.; Patras, M.A.; Kurz, M.; Zubeil, F.; Förster, F.; Schuler, S.; Bauer, A.; Hammann, P.; Vilcinskas, A.; Schäberle, T.F.; et al. Molecular networking-guided discovery and characterization of stechlisins, a group of cyclic lipopeptides from a *Pseudomonas* sp. *J. Nat. Prod.* **2020**, *83*, 2607–2617. [\[CrossRef\]](#) [\[PubMed\]](#)
111. Um, S.; Seibel, E.; Schalk, F.; Balluff, S.; Beemelmans, C. Targeted isolation of saalfelduracin B-D from *Amycolatopsis saalfeldensis* using LC-MS/MS-based molecular networking. *J. Nat. Prod.* **2021**, *84*, 1002–1011. [\[CrossRef\]](#) [\[PubMed\]](#)
112. Guo, J.; Huan, T. Comparison of full-scan, data-dependent, and data-independent acquisition modes in liquid chromatography-mass spectrometry based untargeted metabolomics. *Anal. Chem.* **2020**, *92*, 8072–8080. [\[CrossRef\]](#)
113. Yu, F.C.; Teo, G.C.; Kong, A.T.; Fröhlich, K.; Li, G.X.; Demichev, V.; Nesvizhskii, A.I. Analysis of DIA proteomics data using MSFragger-DIA and FragPipe computational platform. *Nat. Commun.* **2023**, *14*, 4154. [\[CrossRef\]](#)
114. Chang, S.S.; Li, Y.H.; Huang, X.Y.; He, N.; Wang, M.Y.; Wang, J.H.; Luo, M.N.; Li, Y.; Xie, Y.Y. Bioactivity-based molecular networking-guided isolation of epicolidines A-C from the endophytic fungus *Epicoccum* sp. 1-042. *J. Nat. Prod.* **2024**, *87*, 1582–1590. [\[CrossRef\]](#)
115. Damiani, T.; Jarmusch, A.K.; Aron, A.T.; Petras, D.; Phelan, V.V.; Zhao, H.N.; Bittremieux, W.; Acharya, D.D.; Ahmed, M.M.A.; Bauermeister, A.; et al. A universal language for finding mass spectrometry data patterns. *Nat. Methods* **2025**, *22*, 1247–1254. [\[CrossRef\]](#)
116. Berger, T.; Alenfelder, J.; Steinmüller, S.; Heimann, D.; Gohain, N.; Petras, D.; Wang, M.X.; Berger, R.; Kostenis, E.; Reher, R. A massQL-integrated molecular networking approach for the discovery and substructure annotation of bioactive cyclic peptides. *J. Nat. Prod.* **2024**, *87*, 692–704. [\[CrossRef\]](#)
117. Hou, X.M.; Li, Y.Y.; Shi, Y.W.; Fang, Y.W.; Chao, R.; Gu, Y.C.; Wang, C.Y.; Shao, C.L. Integrating molecular networking and ^1H NMR to target the isolation of chrysogeamides from a library of marine-derived *Penicillium* fungi. *J. Org. Chem.* **2019**, *84*, 1228–1237. [\[CrossRef\]](#)
118. Flores-Bocanegra, L.; Al Subeh, Z.Y.; Egan, J.M.; El-Elmat, T.; Raja, H.A.; Burdette, J.E.; Pearce, C.J.; Linington, R.G.; Oberlies, N.H. Dereplication of fungal metabolites by NMR-based compound networking using MADByTE. *J. Nat. Prod.* **2022**, *85*, 614–624. [\[CrossRef\]](#)
119. Borges, R.M.; Ferreira, G.D.; Campos, M.M.; Teixeira, A.M.; Costa, F.D.; das Chagas, F.O.; Colonna, M. NMR as a tool for compound identification in mixtures. *Phytochem. Anal.* **2023**, *34*, 385–392. [\[CrossRef\]](#)
120. Egan, J.M.; van Santen, J.A.; Liu, D.Y.; Linington, R.G. Development of an NMR-based platform for the direct structural annotation of complex natural products mixtures. *J. Nat. Prod.* **2021**, *84*, 1044–1055. [\[CrossRef\]](#)
121. 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. [\[CrossRef\]](#)
122. Agrawal, P.; Amir, S.; Deepak; Barua, D.; Mohanty, D. RiPPMiner-Genome: A web resource for automated prediction of crosslinked chemical structures of RiPPs by genome mining. *J. Mol. Biol.* **2021**, *433*, 166887. [\[CrossRef\]](#)
123. Mahmud, T. Isotope tracer investigations of natural products biosynthesis: The discovery of novel metabolic pathways. *J. Label. Compd. Radiopharm.* **2007**, *50*, 1039–1051. [\[CrossRef\]](#)
124. Saad, H.; Aziz, S.; Gehringer, M.; Kramer, M.; Straetener, J.; Berscheid, A.; Brötz-Oesterheld, H.; Gross, H. Nocathioamides, Uncovered by a tunable metabologenomic approach, define a novel class of chimeric lanthipeptides. *Angew. Chem.-Int. Ed.* **2021**, *60*, 16472–16479. [\[CrossRef\]](#)

125. Reher, R.; Kim, H.W.; Zhang, C.; Mao, H.H.; Wang, M.X.; 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](#)]
126. Sun, Z.L.; Wu, M.Y.; Zhong, B.Y.; Wu, J.S.; Liu, D.; Ren, J.W.; Fan, S.L.; Lin, W.H.; Fan, A.L. Target discovery of dhilirane-type meroterpenoids by biosynthesis guidance and tailoring enzyme catalysis. *J. Am. Chem. Soc.* **2024**, *146*, 30242–30251. [[CrossRef](#)]
127. Shin, D.; Byun, W.S.; Kang, S.; Kang, I.; Bae, E.S.; An, J.S.; Im, J.H.; Park, J.; Kim, E.; Ko, K.; et al. Targeted and logical discovery of piperazic acid-bearing natural products based on genomic and spectroscopic signatures. *J. Am. Chem. Soc.* **2023**, *145*, 19676–19690. [[CrossRef](#)]
128. Morgan, K.D.; Williams, D.E.; Patrick, B.O.; Remigy, M.; Banuelos, C.A.; Sadar, M.D.; Ryan, K.S.; Andersen, R.J. Incarnatapeptins A and B, nonribosomal peptides discovered using genome mining and ^1H ^{15}N HSQC-TOCSY. *Org. Lett.* **2020**, *22*, 4053–4057. [[CrossRef](#)]
129. Huang, H.M.; Yue, L.G.; Deng, F.Y.; Wang, X.Y.; Wang, N.; Chen, H.; Li, H.Y. NMR-metabolomic profiling and genome mining drive the discovery of cyclic decapeptides from a marine *Streptomyces*. *J. Nat. Prod.* **2023**, *86*, 2122–2130. [[CrossRef](#)]
130. Park, J.; Shin, Y.H.; Hwang, S.; Kim, J.; Moon, D.H.; Kang, I.; Ko, Y.J.; Chung, B.; Nam, H.; Kim, S.; et al. Discovery of terminal oxazole-bearing natural products by a targeted metabologenomic approach. *Angew. Chem.-Int. Ed.* **2024**, *63*, e202402465. [[CrossRef](#)]
131. Cox, C.L.; Tietz, J.I.; Melby, J.O.; Sokolowski, K.; Doroghazi, J.R.; Mitchell, D.A. Nucleophilic 1,4-additions for natural product discovery. *Abstr. Gen. Meet. Am. Soc. Microbiol.* **2014**, *114*, 2438. [[CrossRef](#)] [[PubMed](#)]
132. McCaughey, C.S.; van Santen, J.A.; van der Hooft, J.J.J.; Medema, M.H.; Linington, R.G. An isotopic labeling approach linking natural products with biosynthetic gene clusters. *Nat. Chem. Biol.* **2022**, *18*, 295–304. [[CrossRef](#)] [[PubMed](#)]
133. Fergusson, C.H.; Saulog, J.; Paulo, B.S.; Wilson, D.M.; Liu, D.Y.; Morehouse, N.J.; Waterworth, S.; Barkei, J.; Gray, C.A.; Kwan, J.C.; et al. Discovery of a lagriamide polyketide by integrated genome mining, isotopic labeling, and untargeted metabolomics. *Chem. Sci.* **2024**, *15*, 8089–8096. [[CrossRef](#)]
134. Chen, D.W.; Song, Z.J.; Han, J.J.; Liu, J.M.; Liu, H.W.; Dai, J.G. Targeted discovery of glycosylated natural products by tailoring enzyme-guided genome mining and MS-based metabolome analysis. *J. Am. Chem. Soc.* **2024**, *146*, 9614–9622. [[CrossRef](#)]
135. Lee, J.; Um, S.; Kim, E.H.; Kim, S.H. Genomic and metabolomic analyses of *Nocardiopsis maritima* YSL2 as the mycorrhizosphere bacterium of *Suaeda maritima* (L.) Dumort. *J. Nat. Prod.* **2024**, *87*, 733–742. [[CrossRef](#)]
136. Ahmed, M.M.A.; Boudreau, P.D. LCMS-metabolomic profiling and genome mining of *Delftia lacustris* DSM 21246 revealed lipophilic delftibactin metallophores. *J. Nat. Prod.* **2024**, *87*, 1384–1393. [[CrossRef](#)]
137. Yang, F.; Sang, M.L.; Lu, J.R.; Zhao, H.M.; Zou, Y.K.; Wu, W.; Yu, Y.; Liu, Y.W.; Ma, W.C.; Zhang, Y.; et al. Somalactams A-D: Anti-inflammatory macrolide lactams with unique ring systems from an arctic Actinomycete Strain. *Angew. Chem.-Int. Ed.* **2023**, *62*, e202218085. [[CrossRef](#)]
138. Tao, H.; Lauterbach, L.; Bian, G.K.; Chen, R.; Hou, A.W.; Mori, T.; Cheng, S.; Hu, B.; Lu, L.; Mu, X.; et al. Discovery of non-squalene triterpenes. *Nature* **2022**, *606*, 414–419. [[CrossRef](#)] [[PubMed](#)]
139. Mullaney, M.W.; Duncan, K.R.; Elsayed, S.S.; Garg, N.; van der Hooft, J.J.J.; Martin, N.I.; Meijer, D.; Terlouw, B.R.; Biermann, F.; Blin, K.; et al. Artificial intelligence for natural product drug discovery. *Nat. Rev. Drug Discov.* **2023**, *22*, 895–916. [[CrossRef](#)] [[PubMed](#)]
140. Xue, H.T.; Stanley-Baker, M.; Kong, A.W.K.; Li, H.L.; Goh, W.W.B. Data considerations for predictive modeling applied to the discovery of bioactive natural products. *Drug Discov. Today* **2022**, *27*, 2235–2243. [[CrossRef](#)] [[PubMed](#)]
141. Schneider, P.; Altmann, K.H.; Schneider, G. Generating bioactive natural product-inspired molecules with machine intelligence. *Chimia* **2022**, *76*, 396–401. [[CrossRef](#)]
142. Arora, S.; Chettri, S.; Percha, V.; Kumar, D.; Latwal, M. Artificial intelligence: A virtual chemist for natural product drug discovery. *J. Biomol. Struct. Dyn.* **2024**, *42*, 3826–3835. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.