Review Article

Microbial Metabolite Databases: A Comprehensive Review of Development, Integration, and Future Directions

Tamil Bharathi Viswanathan and Gnanendra Shanmugam*

Department of Biotechnology, Vivekanandha College of arts and Sciences for Women (Autonomous), Elayampalyam, Tiruchengode, Tamilnadu, India.

*Correspondence: gnani@vicas.org (GS)

Abstract : Microbial metabolites are bioactive compounds produced by microorganisms, which plays a crucial role in biotechnology, pharmaceuticals, and natural product research. The rapid advancements in high-throughput sequencing, mass spectrometry, and NMR spectroscopy have led to an exponential increase in metabolite data making efficient data management essential. This need has driven the development of specialized microbial metabolite databases, designed to facilitate data storage, annotation, and analysis. This review explores the microbial metabolite databases, particularly focusing on their architecture, data curation strategies, and integration of genomic, metabolomic, and spectral data. Furthermore, a comparative analysis of major databases highlights their strengths, limitations, and areas for improvement. Additionally, we discuss the role of computational approaches in metabolite annotation, biosynthetic pathway mapping, and predictive biosynthesis models. Despite it's significant progress, challenges such as data standardization, cross-database interoperability, and reproducibility remain major hurdles. To overcome these issues, recent advancements such as cloud-based repositories, multi-omics integration, and cheminformatics-driven metabolite prediction offers a promising solution. As this research advances, strengthening database connectivity, fostering collaboration, and integrating emerging technologies will be crucial for unlocking the vast potential of microbial metabolite exploration.

Keywords: Microbial metabolites; high-throughput sequencing; cloud-based repositories; chemoinformatics.

Citation: Tamil Bharathi Viswanathan and Gnanendra S. Microbial Metabolite Databases: A Comprehensive Review of Development, Integration, and Future Directions. Int J Adv Interdis Res 2025, e006.

Received: 14 April 2025
Revised: 12 May 2025
Accepted: 16 May 2025
Published: 17 May 2025

Introduction

During metabolic processes, microorganisms produce a bioactive compounds known as microbial metabolites, which plays a crucial role in growth, survival, and interactions with their environment [1,2]. These metabolites are categorized as primary and secondary metabolites. Primary metabolites are synthesized by an organism during the growth phase, and are integral to metabolic pathways, supporting essential for the physiological processes such as growth, development, and reproduction [3]. In contrast, the secondary metabolites, a low molecular weight compounds, with functions beyond primary physiological processes, playing vital roles in defense, competition, and signaling within their ecological niche [4]. Microbial metabolites have a profound impact across various industries, revolutionizing processes and products. These metabolites are crucial in diverse fields due to their unique chemical properties and biological activities [5].

In the pharmaceutical industry, these metabolites act as the basis for

numerous drugs, including antibiotics (e.g., penicillin and rifamycin), anticancer agents (e.g., doxorubicin), and immunosuppressants (e.g., cyclosporine) [6,7]. They have been crucial in drug discovery and development, offering effective treatments for infectious diseases and chronic conditions [8]. In agriculture, microbial metabolites improve crop productivity and protect plants from pests and diseases [9]. They are used as biofertilizers to enhance nutrient availability, biopesticides to control pests, and plant growth promoters, such as auxins and gibberellins, to stimulate plant development [10].

In the food industry, these metabolites enhance food preservation, flavor, and nutritional value [11]. Fermentation processes utilize metabolites like lactic acid to produce yogurt and cheese, while natural preservatives such as nisin prevent spoilage [12]. Additionally, certain metabolites contribute to desirable flavors, such as diacetyl, which imparts a buttery taste to foods [13]. In environmental biotechnology, microbial metabolites play a vital role in bioremediation and biofuel production [14]. Microorganisms degrade toxic pollutants, clean up oil spills using biosurfactants, and produce renewable biofuels like bioethanol and biodiesel [15].

Furthermore, microbial metabolites have extensive applications in industrial biotechnology. They are used to produce enzymes, organic acids, and amino acids for various processes, including the manufacture of detergents, food supplements, and chemical intermediates [16]. Also, they play a crucial role in human health and nutrition. Probiotics, which are live microorganisms that promote gut health, and prebiotics, which support beneficial gut bacteria, are essential for maintaining a healthy digestive system [17]. Microbial-derived vitamins like B12 and riboflavin also contribute significantly to human nutrition [18].

The vast diversity of microbial metabolites offers numerous benefits across multiple fields. Their applications in pharmaceuticals, agriculture, and environmental biotechnology make them indispensable for human welfare. With advances in metabolomics and microbial research, their potential for future innovations is immense. With this rapid progress, the volume of microbial metabolite data has grown exponentially. This surge highlights the need for reliable databases to manage, organize, and share information effectively. Such databases serve as centralized platforms for storing chemical structures, bioactivity data, and biosynthetic pathways, enabling seamless data access and analysis. They provide a fundamental framework for research acceleration, innovation, and discovery across drug development, agriculture, and biotechnology.

Evolution of Microbial Metabolite Research

Microbial metabolite research has significantly evolved over the past century, primarily due to advances in analytical and sequencing technologies. Initially, the discovery of bioactive microbial compounds was largely dependent on culture-based techniques, limiting the exploration of microbial diversity [19]. However, with the advent of genomic sequencing, metagenomics, and

metabolomics, researchers can now explore a much broader spectrum of microbial metabolites, elucidating many that have long remained undiscovered [20]. These advancements have not only expanded the diversity of microbial metabolites but also enhanced the ability to understand their biological roles and biosynthetic pathways.

Historical Advancements in Microbial Metabolomics

The roots of microbial metabolite research can be traced back to the discovery of penicillin by Alexander Fleming in 1928, marking a turning point in antimicrobial drug development [21]. This milestone led to the systematic screening of soil microbes, particularly Actinobacteria, which resulted in the discovery of streptomycin, erythromycin, tetracycline and others [22, 23]. During the mid-20th century, solvent extraction and chromatography techniques, such as thin-layer chromatography (TLC) and high-performance liquid chromatography (HPLC), were widely used to isolate and analyze microbial metabolites [24].

The genomic era brought a paradigm shift in microbial metabolite research, where the sequencing of microbial genomes provided deeper insights into biosynthetic gene clusters (BGCs), which encode the machinery for metabolite production [25]. This led to the development of genome mining approaches, allowing researchers to predict and identify novel secondary metabolites computationally rather than relying solely on bioactivity-guided screening [26]. The launch of large-scale initiatives like the Human Microbiome Project (HMP) in 2006 further propelled microbial metabolite research, highlighting the significance of microbiota-derived metabolites in human health and disease [27].

Impact of High-Throughput Sequencing and Analytical Techniques

The emergence of next-generation sequencing (NGS) and third-generation sequencing (TGS) technologies, such as Illumina, PacBio, and Oxford Nanopore, has advanced the study of microbial metabolites by providing more comprehensive genomic insights [28]. These technologies enable detailed metagenomic and metatranscriptomic analyses, allowing scientists to reconstruct microbial communities and associate them with metabolite production, even in unculturable organisms [29].

In parallel, a significant progress has been made in analytical chemistry techniques, where the integration of ultra-high-performance liquid chromatography (UHPLC), high-resolution mass spectrometry (HR-MS), and nuclear magnetic resonance (NMR) spectroscopy has greatly enhanced the detection, quantification, and structural elucidation of microbial metabolites [30]. Moreover, tools like Global Natural Products Social Molecular Networking (GNPS) and molecular networking algorithms have improved metabolite identification by comparing spectral data across large datasets [31]. These advancements have not only increased the efficiency of metabolite annotation but have also facilitated large-scale data sharing within the scientific

community.

Shift from Culture-Dependent to Culture-Independent Metabolomics

Historically, microbial metabolite discovery relied on culture-dependent methods, which, despite being effective, presented a major limitation—more than 99% of microbial species remain unculturable under standard laboratory conditions [32]. This restriction meant that the vast majority of microbial metabolites remained undiscovered. To overcome this, culture-independent approaches, such as metagenomics, metatranscriptomics, and single-cell metabolomics, have been widely adopted [33].

Metagenomics has emerged as a powerful tool for exploring the untapped biosynthetic potential within environmental and host-associated microbiomes [34]. By assembling microbial genomes directly from environmental samples, researchers can predict and characterize biosynthetic pathways without requiring pure cultures. Metatranscriptomics further refines this approach by providing insights into actively expressed genes, helping differentiate between dormant and functionally active biosynthetic pathways [35].

Another breakthrough has been Mass Spectrometry Imaging (MSI), which allows for the spatial mapping of metabolite production within microbial communities. Coupled with stable isotope labeling techniques, MSI has provided a clearer understanding of microbial interactions and metabolic exchanges within natural ecosystems [36].

Together, these advancements have transformed microbial metabolomics into a highly interdisciplinary field, integrating genomics, bioinformatics, and analytical chemistry to discover and characterize novel microbial metabolites. As sequencing and analytical technologies continue to improve, researchers will gain even deeper insights into microbial metabolic diversity, opening new avenues for drug discovery, biotechnology, and personalized medicine.

Microbial Metabolite Database

Microbial metabolite databases serve as essential repositories for storing, organizing, and analyzing metabolite-related data, facilitating drug discovery, metabolomics, and biochemical pathway research. The increasing availability of high-throughput sequencing and metabolomics technologies has led to the development of multiple databases, each designed for different applications, including chemical structure annotation, biosynthetic gene cluster (BGC) prediction, and pathway reconstruction This review provides a comparative analysis of key microbial metabolite databases, focusing on their data structure, curation methodologies, accessibility, and analytical capabilities.

Classification of Microbial Metabolite Databases

Microbial metabolite databases can be broadly categorized into chemical-centric, genome-centric, and pathway-centric databases, each offering distinct functionalities.

Chemical-Centric Databases

These databases primarily stores structural, physicochemical, and spectral data of microbial metabolites, facilitating compound identification and dereplication. NPAtlas, for instance, contains over 30,000 natural products, providing taxonomic details, molecular weights, and bioactivity data [37]. Similarly, MetaboLights serves as a repository for metabolomics experiments, integrating raw spectral data, metadata, and experimental protocols [38].

Genome-Centric Databases

With advancements in genome mining, databases focusing on BGCs and gene-metabolite relationships have gained prominence. MIBiG provides manually curated BGC information, linking genomic sequences to experimentally validated metabolites [39]. antiSMASH, on the other hand, predicts putative BGCs using computational algorithms, offering comparative analysis with known clusters [40]. These databases enable researchers to identify novel bioactive compounds directly from genomic data, bypassing traditional culture-based approaches.

Pathway-Centric Databases

Pathway-focused databases, such as MetaCyc and KEGG, store information on metabolic and biosynthetic pathways, allowing researchers to reconstruct biochemical networks and understand microbial metabolism [41]. These databases integrate gene, enzyme, and metabolite interactions, providing a systems biology perspective on microbial metabolite production.

Comparative Assessment of Major Databases

The differences in data structure, curation, and accessibility significantly impact the usability of these databases for various research applications. A comparative analysis of widely used microbial metabolite databases is summarized in Table 1.

Challenges and Limitations in Database Utilization

Despite their utility, microbial metabolite databases face several limitations that hinder their effectiveness in large-scale metabolomics research.

Data Fragmentation and lack of standardization

The absence of a unified data-sharing framework results in inconsistencies across different databases. For example, chemical-centric databases (e.g., NPAtlas, MetaboLights) do not always integrate genomic information, while genome-centric databases (e.g., MIBiG, antiSMASH) lack detailed metabolite characterizations.

Curation Bottlenecks

While some databases rely on manual curation (e.g., MIBiG, MetaCyc), ensuring high-quality data, others use automated pipelines (e.g., antiSMASH), which may introduce annotation errors [42]. Inconsistent metadata submissions further impact database reliability [43].

Table 1. Comparative Analysis of Microbial Metabolite Databases

Database	Primary Focus	Data Coverage	Curation	Accessibility	Key Features
NPAtlas	Natural product chemistry	Chemical structures, taxonomic data	Manually curated	Open-access	Structure-based search
MetaboLights	Metabolomics data repository	Spectral data, metadata	Curated by submitters	Open-access	Experimental metadata integration
MIBiG	Biosynthetic gene clusters	Gene-metabolite links, BGCs	Expert manual curation	Open-access	Genome-metabolite correlation
antiSMASH	Genome mining for BGCs	BGC prediction, genome annotation	Automated	Open-access	Comparative BGC analysis
MetaCyc	Metabolic pathways	Enzyme, metabolite, pathway data	Expert curation	Open-access	Metabolic network visualization
KEGG	Pathway-based interactions	Multi-omics integration (genes, proteins, metabolites)	Curated by KEGG team	Subscription-based	Pathway mapping tools

Access Restrictions and Usability Constraints

Open-access databases facilitate global collaboration, whereas some repositories, such as KEGG, require subscription fees, limiting accessibility [44]. Additionally, the lack of user-friendly query interfaces in some databases complicates large-scale data retrieval.

Biosynthetic Pathway Mapping and Functional Insights

Microbial metabolites are synthesized through intricate biosynthetic pathways, often encoded within biosynthetic gene clusters (BGCs). These pathways involve multi-step enzymatic reactions that convert primary metabolites into structurally diverse secondary metabolites, such as antibiotics, antifungals, and anticancer agents [45,46]. The identification and mapping of these pathways are crucial for understanding microbial metabolism and engineering novel bioactive compounds. Advancements in sequencing technologies, genome mining, and metabolic modeling have significantly improved the ability to reconstruct these pathways. Traditional culture-based approaches have now been supplemented by bioinformatics tools like antiSMASH, PRISM, and BiG-FAM, which enable BGC detection, functional annotation, and comparative analysis [47,48].

Biosynthetic pathway mapping not only aids in metabolite discovery but also provides insights into evolutionary relationships among microbial species. Comparative genomics has revealed that horizontal gene transfer plays a significant role in shaping BGC diversity, leading to the emergence of novel metabolites [49]. Additionally, functional validation techniques such as CRISPR-Cas9 genome editing, isotopic labeling, and heterologous expression allow researchers to confirm enzymatic functions and regulatory mechanisms within these pathways [50,51]. Moreover, metabolic engineering strategies have been employed to optimize pathway flux and enhance metabolite yield,

particularly in industrial biotechnology and pharmaceutical production [52]. Despite these advancements, challenges remain in activating silent BGCs, resolving complex regulatory networks, and integrating multi-omics data for complete pathway elucidation [53]. Future developments in AI-driven biosynthetic modeling and synthetic biology will further enhance the predictive power and application of biosynthetic pathway mapping.

Challenges in Microbial Metabolite Data Curation and Standardization

The rapid expansion of microbial metabolomics has led to an overwhelming influx of metabolite-related data, necessitating systematic curation and standardization to ensure accuracy, accessibility, and reproducibility. However, several challenges hinder the seamless integration of microbial metabolite data across different databases and analytical platforms.

One of the primary challenges is the heterogeneity of data sources, as microbial metabolites are identified using diverse techniques, including mass spectrometry (MS), nuclear magnetic resonance (NMR), and chromatography-based approaches [54]. Variability in instrumentation, sample preparation, and spectral analysis can lead to inconsistencies in reported metabolite structures and concentrations. Additionally, many metabolites lack well-defined reference standards, making cross-platform validation difficult [55].

Another significant issue is data annotation and nomenclature inconsistencies. Different databases, such as MIBiG, NPAtlas, and MetaboLights, often employ varying formats for storing chemical structures, bioactivities, and metadata [56]. The absence of a universal metabolite ontology complicates data comparison and integration across repositories [57]. Moreover, the misclassification of metabolites due to insufficient spectral resolution or ambiguous structural features further complicates annotation efforts [56].

Data curation bottlenecks also arise due to the manual nature of metabolite validation, requiring expert review to eliminate errors and redundancies [58]. However, automated curation pipelines leveraging machine learning and artificial intelligence are emerging as promising solutions to enhance data quality and reduce human error [59].

Standardization efforts, such as the Metabolomics Standards Initiative (MSI) and FAIR (Findable, Accessible, Interoperable, Reusable) data principles, aim to improve data reproducibility and sharing [60,61]. However, widespread adoption remains a challenge due to incompatibilities in legacy databases, proprietary data formats, and insufficient metadata reporting [62]. Moving forward, greater collaboration between bioinformaticians, experimental biologists, and database curators is essential to establish robust frameworks for metabolite data standardization and interoperability [63].

Advances in Metabolomics Technology and Data Storage

Metabolomics has rapidly evolved with advancements in analytical

technologies and data management strategies, enabling a deeper understanding of biochemical processes at a systems level. The integration of high-resolution mass spectrometry (HR-MS) and nuclear magnetic resonance (NMR) spectroscopy has significantly improved metabolite detection, identification, and quantification. HR-MS provides high sensitivity and mass accuracy, making it suitable for untargeted metabolomic studies, whereas NMR spectroscopy offers robust and reproducible structural elucidation of metabolites without sample destruction [64]. The combination of these techniques enhances data accuracy and expands metabolite coverage, facilitating comprehensive metabolic profiling across various biological systems.

With the increasing complexity and volume of metabolomics data, the need for efficient data storage and sharing platforms has become paramount. Cloud-based repositories and open-access databases have emerged as essential tools for researchers, allowing seamless data integration, accessibility, and collaboration [65]. Platforms such as MetaboLights, GNPS, and Metabolomics Workbench provide standardized data formats and interoperability, ensuring reproducibility and transparency in metabolomics studies [66]. These repositories enable global sharing of metabolomics datasets, fostering collaborative research and comparative analyses across different laboratories and disciplines.

Furthermore, the advent of multi-omics approaches has revolutionized metabolite research by integrating metabolomics with genomics, transcriptomics, and proteomics. This holistic strategy provides deeper insights into metabolic pathways, disease mechanisms, and biomarker discovery. By correlating metabolic changes with genetic and proteomic alterations, researchers can uncover novel regulatory networks and potential therapeutic targets. Advances in bioinformatics and machine learning further enhance the integration of multi-omics data, enabling predictive modeling and personalized medicine applications. As metabolomics continues to advance, the synergy between cutting-edge analytical technologies, cloud-based data management, and multi-omics approaches will play a pivotal role in driving innovation and translational research in the biomedical and life sciences fields.

Limitations and Future Directions

Despite significant progress, several challenges remain in metabolomics research. One major limitation is the incompleteness of microbial metabolite databases. Many existing databases lack comprehensive annotations and coverage of microbial metabolites, which restricts the accurate identification of novel compounds and their biological significance. Expanding and standardizing microbial metabolite libraries will be crucial for enhancing our understanding of microbial metabolism and its implications in health and disease.

Another limitation is the variability in data acquisition, processing, and

interpretation, which complicates cross-laboratory comparisons. Differences in sample preparation, instrumentation, and analytical protocols can introduce inconsistencies in metabolomics datasets. Standardized methodologies and robust quality control measures must be implemented to ensure reproducibility and comparability across studies.

Additionally, the lack of standardized methodologies for data integration and cross-database interoperability poses a challenge in metabolomics research. Differences in data formats, processing pipelines, and analytical techniques create inconsistencies that hinder effective data sharing and comparative studies. Developing universal data standards, improving metadata annotation, and implementing automated curation pipelines are necessary to overcome these barriers and enhance the reproducibility of metabolomics findings.

Future advancements in metabolomics are expected to focus on improving natural product discovery, refining data integration strategies, and expanding computational approaches. The application of artificial intelligence and machine learning in metabolomics is anticipated to enhance metabolite classification, metabolic pathway reconstruction, and predictive modeling. Furthermore, integrating metabolomics with metagenomics and transcriptomics will provide deeper insights into host-microbe interactions and their metabolic functions. The adoption of real-time metabolomics analysis, supported by miniaturized and portable analytical devices, will further revolutionize point-of-care diagnostics and environmental monitoring. Addressing these limitations and leveraging emerging technologies will be essential in maximizing the impact of metabolomics in biomedical research, environmental studies, and drug discovery.

Conclusion

This review provides a comprehensive analysis of microbial metabolite databases, tracing their evolution, advancements, and future directions. It delves into database architecture, focusing on data curation, cheminformatics, and bioinformatics integration for efficient metabolite annotation and biosynthetic pathway mapping. The challenges of data standardization, reproducibility, and cross-database interoperability are critically examined, highlighting their impact on research reliability. Additionally, it explores emerging trends such as cloud-based repositories, multi-omics integration, and predictive biosynthetic modeling. By addressing current limitations and future prospects, this review underscores the pivotal role of microbial metabolite databases in biotechnology, pharmaceuticals, and natural product discovery.

References

[1] Berdy J. Bioactive microbial metabolites. The Journal of antibiotics. 2005 Jan;58(1):1-26.

[2] Hashem ZS. Bacterial Metabolites in Defense: A Crucial Aspect of Microbial Interaction and Host Protection. InMetabolic Dynamics in Host-Microbe Interaction 2025

Feb 6 (pp. 101-120). Singapore: Springer Nature Singapore.

[3] Saravanan C. Primary and secondary microbial metabolites. Industrial microbiology.:24.

[4] Al-Khayri JM, Rashmi R, Toppo V, Chole PB, Banadka A, Sudheer WN, Nagella P, Shehata WF, Al-Mssallem MQ, Alessa FM, Almaghasla MI. Plant secondary metabolites: The weapons for biotic stress management. Metabolites. 2023 May 31;13(6):716.

[5] Davies J. Specialized microbial metabolites: functions and origins. The Journal of antibiotics. 2013 Jul;66(7):361-4.

[6] Ramírez-Rendon D, Passari AK, Ruiz-Villafán B, Rodríguez-Sanoja R, Sánchez S, Demain AL. Impact of novel microbial secondary metabolites on the pharma industry. Applied Microbiology and Biotechnology. 2022 Mar;106(5):1855-78.

[7] Vaishnav P, Demain AL. Unexpected applications of secondary metabolites. Biotechnology Advances. 2011 Mar 1;29(2):223-9.

[8] Saha S, Rajpal DK, Brown JR. Human microbial metabolites as a source of new drugs. Drug discovery today. 2016 Apr 1;21(4):692-8.

[9] Jayaprakashvel M, Mathivanan N. Management of plant diseases by microbial metabolites. Bacteria in agrobiology: plant nutrient management. 2011:237-65.

[10] Pathma J, Kennedy RK, Bhushan LS, Shankar BK, Thakur K. Microbial biofertilizers and biopesticides: nature's assets fostering sustainable agriculture. Recent developments in microbial technologies. 2021:39-69.

[11] Singh R, Kumar M, Mittal A, Mehta PK. Microbial metabolites in nutrition, healthcare and agriculture. 3 Biotech. 2017 May;7:1-4.

[12] Ibrahim SA, Ayivi RD, Zimmerman T, Siddiqui SA, Altemimi AB, Fidan H, Esatbeyoglu T, Bakhshayesh RV. Lactic acid bacteria as antimicrobial agents: Food safety and microbial food spoilage prevention. Foods. 2021 Dec 17;10(12):3131.

[13] Cheetham PS. Natural sources of flavours. Food flavour technology. 2010 Feb 8:127-77.

[14] Dangi AK, Sharma B, Hill RT, Shukla P. Bioremediation through microbes: systems biology and metabolic engineering approach. Critical reviews in biotechnology. 2019 Jan 2;39(1):79-98.

[15] Oyetunji OE, Kotun BC, Thonda OA, Ademola EA. The role of biosurfactants in biofuel production. InAdvancements in biosurfactants research 2023 Jan 26 (pp. 371-395). Cham: Springer International Publishing.

[16] Bhalla TC, Sharma NN, Sharma M. Production of metabolites, industrial enzymes, amino acid, organic acids, antibiotics, vitamins and single cell proteins.

[17] Quigley EM. Prebiotics and probiotics in digestive health. Clinical Gastroenterology and Hepatology. 2019 Jan 1;17(2):333-44.

[18] Averianova LA, Balabanova LA, Son OM, Podvolotskaya AB, Tekutyeva LA. Production of vitamin B2 (riboflavin) by microorganisms: an overview. Frontiers in Bioengineering and Biotechnology. 2020 Nov 12;8:570828.

[19] Rocha-Martin J, Harrington C, Dobson AD, O'Gara F. Emerging strategies and integrated systems microbiology technologies for biodiscovery of marine bioactive compounds. Marine drugs. 2014 Jun 10;12(6):3516-59.

[20] Shen B, Tang Y, Baltz RH, Gonzalez R. Introduction to the special issue:"Natural Product Discovery and Development in the Genomic Era: 2021". Journal of Industrial Microbiology and Biotechnology. 2021 Apr;48(3-4):kuab030.

[21] Fleming A. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. British journal of experimental pathology. 1929 Jun;10(3):226.

[22] Waksman SA, Katz E, Vining LC. Nomenclature of the actinomycins. Proceedings of the National Academy of Sciences. 1958 Jun;44(6):602-12.

[23] Quinn GA, Banat AM, Abdelhameed AM, Banat IM. Streptomyces from traditional medicine: sources of new innovations in antibiotic discovery. Journal of medical microbiology. 2020 Aug;69(8):1040-8.

[24] Kota MS, Gavaskar D, Mondal MD, Somanathan T. Chemistry of natural products. Jec publication.

[25] Ziemert N, Alanjary M, Weber T. The evolution of genome mining in microbes–a review. Natural product reports. 2016;33(8):988-1005.

[26] Ziemert N, Weber T, Medema MH. 6.03-Genome mining approaches to bacterial natural product discovery. Liu HW (Ben), Begley TP (editors). Comprehensive Natural Products III. Oxford: Elsevier. 2020:19-33.

[27] Gevers D, Knight R, Petrosino JF, Huang K, McGuire AL, Birren BW, Nelson KE, White O, Methé BA, Huttenhower C. The Human Microbiome Project: a community resource for the healthy human microbiome.

[28] Akacin I, Ersoy Ş, Doluca O, Güngörmüşler M. Comparing the significance of the utilization of next generation and third generation sequencing technologies in microbial metagenomics. Microbiological Research. 2022 Nov 1;264:127154.

[29] Verma A, Sharma A, Kumar M, Bansal S, Shrivastava R. Techniques and challenges in studies related with human gut microbiome. InHuman-gut microbiome 2022 Jan 1 (pp. 37-57). Academic Press.

[30] Rehm K, Vollenweider V, Kümmerli R, Bigler L. A comprehensive method to elucidate pyoverdines produced by fluorescent Pseudomonas spp. by UHPLC-HR-MS/MS. Analytical and Bioanalytical Chemistry. 2022 Mar;414(8):2671-85.

[31] Zhang M, Otsuki K, Li W. Molecular networking as a natural products discovery strategy. Acta Materia Medica. 2023 Apr 25;2(2):126-41.

[32] Stewart EJ. Growing unculturable bacteria. Journal of bacteriology. 2012 Aug 15;194(16):4151-60.

[33] Wei LQ, Cheong IH, Yang GH, Li XG, Kozlakidis Z, Ding L, Liu NN, Wang H. The application of high-throughput technologies for the study of microbiome and cancer. Frontiers in Genetics. 2021 Jul 28;12:699793.

[34] Alteio LV, Schulz F, Seshadri R, Varghese N, Rodriguez-Reillo W, Ryan E, Goudeau D, Eichorst SA, Malmstrom RR, Bowers RM, Katz LA. Complementary metagenomic approaches improve reconstruction of microbial diversity in a forest soil. Msystems. 2020 Apr 28;5(2):10-128.

[35] Li MM, White RR, Guan LL, Harthan L, Hanigan MD. Metatranscriptomic analyses reveal ruminal pH regulates fiber degradation and fermentation by shifting the microbial community and gene expression of carbohydrate-active enzymes. Animal microbiome. 2021 Apr 23;3(1):32.

[36] Dunham SJ, Ellis JF, Li B, Sweedler JV. Mass spectrometry imaging of complex microbial communities. Accounts of chemical research. 2016 Dec 21;50(1):96-104.

[37] Van Santen JA, Jacob G, Singh AL, Aniebok V, Balunas MJ, Bunsko D, Neto FC, Castaño-Espriu L, Chang C, Clark TN, Cleary Little JL. The natural products atlas: an open access knowledge base for microbial natural products discovery. ACS central science. 2019 Nov 14;5(11):1824-33.

[38] Kale NS, Haug K, Conesa P, Jayseelan K, Moreno P, Rocca-Serra P, Nainala VC, Spicer RA, Williams M, Li X, Salek RM. MetaboLights: an open-access database repository for metabolomics data. Current protocols in bioinformatics. 2016 Mar;53(1):14-3.

[39] Kautsar SA, Blin K, Shaw S, Navarro-Muñoz JC, Terlouw BR, Van Der Hooft JJ, Van Santen JA, Tracanna V, Suarez Duran HG, Pascal Andreu V, Selem-Mojica N. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. Nucleic acids research. 2020 Jan 8;48(D1):D454-8.

[40] Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic acids research. 2019 Jul 2;47(W1):W81-7.

[41] Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, Ong WK, Paley S, Subhraveti P, Karp PD. The MetaCyc database of metabolic pathways and enzymes-a 2019 update. Nucleic acids research. 2020 Jan 8;48(D1):D445-53.

[42] Blin K, Kim HU, Medema MH, Weber T. Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters. Briefings in Bioinformatics. 2019 Jul;20(4):1103-13.

[43] Shi J, Nason M, Tullney M, Alperin JP. Identifying Metadata Quality Issues Across Cultures. College & Research Libraries. 2025 Jan 7;86(1):101.

[44] Reiser L, Berardini TZ, Li D, Muller R, Strait EM, Li Q, Mezheritsky Y, Vetushko A, Huala E. Sustainable funding for biocuration: The Arabidopsis Information Resource (TAIR) as a case study of a subscription-based funding model. Database. 2016 Mar 17;2016:baw018.

[45] Bills GF, Gloer JB. Biologically active secondary metabolites from the fungi. Microbiology spectrum. 2016 Dec 30;4(6):10-128.

[46] Vaishnav P, Demain AL. Unexpected applications of secondary metabolites. Biotechnology Advances. 2011 Mar 1;29(2):223-9.

[47] Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic acids research. 2019 Jul 2;47(W1):W81-7.

[48] Kautsar SA, Blin K, Shaw S, Weber T, Medema MH. BiG-FAM: the biosynthetic gene cluster families database. Nucleic acids research. 2021 Jan 8;49(D1):D490-7.

[49] ZZiemert N, Alanjary M, Weber T. The evolution of genome mining in microbes–a review. Natural product reports. 2016;33(8):988-1005.

[50] Luo Y, Enghiad B, Zhao H. New tools for reconstruction and heterologous expression of natural product biosynthetic gene clusters. Natural product reports. 2016;33(2):174-82.

[51] TTong Y, Charusanti P, Zhang L, Weber T, Lee SY. CRISPR-Cas9 based engineering of actinomycetal genomes. ACS synthetic biology. 2015 Sep 18;4(9):1020-9.

[52] Pickens LB, Tang Y, Chooi YH. Metabolic engineering for the production of natural products. Annual review of chemical and biomolecular engineering. 2011 Jul 15;2(1):211-36.

[53] Cimermancic P, Medema MH, Claesen J, Kurita K, Brown LC, Mavrommatis K, Pati A, Godfrey PA, Koehrsen M, Clardy J, Birren BW. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell. 2014 Jul 17;158(2):412-21.

[54] Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and

towards mechanisms. Nature reviews Molecular cell biology. 2016 Jul;17(7):451-9.

[55] Dorrestein PC, Mazmanian SK, Knight R. Finding the missing links among metabolites, microbes, and the host. Immunity. 2014 Jun 19;40(6):824-32.

[56] Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, Sajed T, Johnson D, Li C, Karu N, Sayeeda Z. HMDB 4.0: the human metabolome database for 2018. Nucleic acids research. 2018 Jan 4;46(D1):D608-17.

[57] Salek RM, Neumann S, Schober D, Hummel J, Billiau K, Kopka J, Correa E, Reijmers T, Rosato A, Tenori L, Turano P. COordination of Standards in MetabOlomicS (COSMOS): facilitating integrated metabolomics data access. Metabolomics. 2015 Dec;11:1587-97.

[58] Stanstrup J, Broeckling CD, Helmus R, Hoffmann N, Mathé E, Naake T, Nicolotti L, Peters K, Rainer J, Salek RM, Schulze T. The metaRbolomics Toolbox in Bioconductor and beyond. Metabolites. 2019 Oct;9(10):200.

[59] Alenezi M, Akour M. AI-Driven Innovations in Software Engineering: A Review of Current Practices and Future Directions. Applied Sciences. 2025 Jan 28;15(3):1344.

[60] Spicer RA, Salek R, Steinbeck C. Compliance with minimum information guidelines in public metabolomics repositories. Scientific data. 2017 Sep 26;4(1):1-8.

[61] Zulfiqar M, Crusoe MR, König-Ries B, Steinbeck C, Peters K, Gadelha L. Implementation of fair practices in computational metabolomics workflows—a case study. Metabolites. 2024 Feb 10;14(2):118.

[62] Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA. Untargeted metabolomics strategies—challenges and emerging directions. Journal of the American Society for Mass Spectrometry. 2016 Sep 13;27(12):1897-905.

[63] Steinbeck C, Conesa P, Haug K, Mahendraker T, Williams M, Maguire E, Rocca-Serra P, Sansone SA, Salek RM, Griffin JL. MetaboLights: towards a new COSMOS of metabolomics data management. Metabolomics. 2012 Oct;8:757-60.

[64] Letertre MP, Dervilly G, Giraudeau P. Combined nuclear magnetic resonance spectroscopy and mass spectrometry approaches for metabolomics. Analytical chemistry. 2020 Nov 6;93(1):500-18.

[65] Perez-Riverol Y, Bai M, da Veiga Leprevost F, Squizzato S, Park YM, Haug K, Carroll AJ, Spalding D, Paschall J, Wang M, Del-Toro N. Discovering and linking public omics data sets using the Omics Discovery Index. Nature biotechnology. 2017 May;35(5):406-9.

[66] Long NP, Nghi TD, Kang YP, Anh NH, Kim HM, Park SK, Kwon SW. Toward a standardized strategy of clinical metabolomics for the advancement of precision medicine. Metabolites. 2020 Jan 29;10(2):51.