https://doi.org/10.33451/florafauna.v32i1pp100-104
FLORA AND FAUNA ISSN 2456 - 9364 (Online)
2026 Vol. 32 No.1 PP 100-104 ISSN 0971 - 6920 (Print)

Genome Mining of Actinomycete Secondary Metabolism Across Biodiversity Hotspots: Tools , Trends and Bioprospecting Insights : A Review

***Sristy Sonar, Tiisungna Helena, Wng Julyni, Olivia Angom, Rushika Ngashepam, Sanatombi Konsam and Laishram Bandana**

Department of Biochemistry,
Manipur University, Indo-Myanmar Road,
MANIPUR- 795003, INDIA
*Corresponding Author
E-mail : sristysonar25@gmail.com

Received **:** 22.04.2026; ***Accepted*** : 08.04.2026

How to cite : Sonar S, Helena T, Julyni W, Angom O, Ngashepham R, Konsam S, Bandana L. Genome Mining of Actinomycete Secondary Metabolism Across Biodiversity Hotspots: Tools , Trends and Bioprospecting Insights : A Review. *Flora and Fauna* 2026. 32(1) : 100-104.

ABSTRACT

Actinomycetes are outstanding generators of natural products crucial for medicine, but significant portions of their biosynthetic repertoire—particularly those in biodiversity hotspots rich with ecological and evolutionary distinctiveness—remain largely unexploited. This review promotes genome mining of actinomycetes from these locales, emphasizing how modern high-resolution sequencing allows for organized revelation of latent secondary-metabolite gene clusters. It commences with a classification of the chief BGC types that fuel actinomycete biochemistry, including NRPS, PKS, RiPPs, and terpenes, accentuating their modular configurations and chemical heterogeneity. Following this, it reviews the bioinformatics arsenal: antiSMASH for BGC localization, BiG-SCAPE for clustering networks, MIBiG for benchmark clusters, and metabolite repositories that support dereplication and cataloging. Biogeographic findings from extensive analyses are then presented, showing that strains from hotspots exhibit specialized BGC inventories, with increased hybrid NRPS-PKS constructs and innovative RiPP categories molded by unique environmental dynamics. The review closes by exploring techniques to pair BGCs with their outputs, delineating evidence hierarchies from genetic alterations and heterologous setups to metabolomics profiling and genomic associations, while highlighting obstacles such as dormant clusters, overlapping pathways, and erroneous spectral alignments. Fusing strategic ecological collection, refined bioinformatics, and multi-omics authentication, it delineates a plan for capturing actinomycete biosynthetic novelty in hotspots to deliver groundbreaking drugs and bioindustrial molecules.

Figure : 00 References : 21 Table : 00

KEY WORDS : Actinomycetes, Biodiversity Hotspots, Biosynthetic gene clusters, Computational tools, Genome mining.

Introduction

Actinomycetes, belonging to the phylogenetically varied Actinobacteria phylum[1], stand out as premier producers of bioactive secondary metabolites[3] in microbial natural products discovery. These Gram-positive bacteria, characterized by high G+C content genomes (ranging 69-78%), are celebrated for generating a vast chemical repertoire, encompassing antibiotics, antifungals, anticancer compounds, immuno suppressants and industrially vital enzymes. More than 70% of clinically approved antibiotics—including streptomycin, tetracycline, erythromycin, and vancomycin—trace their origins to actinomycetes, highlighting their profound influence on contemporary medicine and biotechnology. While Streptomyces, the dominant genus, contributes over 80% of these bioactives, less-explored genera such as Saccharopolyspora, Micromonospora, Actinomadura, and Nocardia offer rich, underexploited scaffolds like enediynes and lanthipeptides. This biosynthetic versatility

ACKNOWLEDGEMENTS : The author(s) are grateful to the HOD, Department of Biochemistry, Manipur University. The authors also would like to acknowledge Dr. Saikat Mukherjee, Assistant Professor S-2, Department of Biochemistry, Manipur University for his immense academic support and access to institutional resources that facilitated this work.

arises from sophisticated modular systems, notably non-ribosomal peptide synthetases (NRPS), polyketide synthases (PKS I, II, III), terpene synthases, and ribosomally synthesized post-translationally modified peptides (RiPPs), all housed in expansive biosynthetic gene clusters (BGCs) spanning 20-150 kb across linear chromosomes and plasmid.[4] Biodiversity hotspots[15], designated by Conservation International for their elevated endemism (>1,500 vascular plant species, >2.5% global flora in <0.5% land area) and vulnerability to degradation, act as critical actinomycete reservoirs. These zones align with microbial hotspots defined by abiotic extremes—nutrient-poor soils, high salinity, thermal variances, metal-laden niches—driving adaptive secondary metabolism. Extremophiles from such sites evolve desiccation-proof polyketides in deserts or salt-tolerant antifungals in mangroves. Geographic barriers and co-evolutionary pressures enhance genetic novelty through horizontal gene transfer (HGT) of BGCs, evident in streptomycete plasmid exchanges. Marine coral ecosystems[5] yield obligate actinomycetes like Salinispora tropica (producer of salinosporamide A, a clinical-stage proteasome inhibitor), while forest litter nurtures endophytes linked to fungal symbionts. Genome mining's necessity stems from flaws in conventional bioassay-driven fractionation, which privileges culturable high-producers (<1% of total diversity, per the "great plate count anomaly"). Culture-independent metagenomics unveils uncultivable actinomycetes in rhizospheres and sediments boasting BGC profiles comparable to isolates. This sequence-centric strategy sidesteps cultivation hurdles, spotlighting orphan BGCs without cognate products. Cornerstone tools like antiSMASH identify >50 BGC classes using hidden Markov models (HMMs), Pfam scans, and gene cassettes, precisely decoding NRPS A-domain substrates and PKS ketosynthase motifs (>90% accuracy). BiG-SCAPE clusters BGCs via alignment-free glocal metrics into families with novelty indices; MIBiG[18] curates >2,000 exemplars for ClusterBlast homology searches; databases like GNPS, NORINE, and StreptomeDB bridge sequences to chemistries. Emerging multi-omics—integrating genomics, transcriptomics, proteomics, metabolomics—illuminates BGC dynamics. India alone boasts 36 biodiversity hotspots (Himalaya subregions, Western Ghats, Indo-Burma fringes, Sundaland extensions, island chains), positioning it as an actinomycete mining powerhouse.

BGC Types and Biosynthetic Logic

Non-Ribosomal Peptide Synthetases (NRPS):

NRPS produce peptides independently of the ribosome, allowing for immense structural diversity, including cyclization and unusual amino acids[21].

Biosynthetic Logic: They operate through a modular assembly line. Each module is responsible for adding one amino acid.

Core Domains:

Adenylation (A) domain: Recognizes and activates specific amino acids.

Thiolation (PCP/T) domain: Carries the growing chain.

Condensation (C) domain: Forms the peptide bond.

Modifications: Tailoring enzymes (methyltransferases, epimerases) are often included.

Polyketide Synthases (PKS):

PKSs produce polyketides[17], including important antibiotics, through a mechanism similar to fatty acid synthesis.

Biosynthetic Logic: Modular (Type I) or Iterative (Type II/III) assembly of carboxylic acid building blocks.

Types:

Type I (Modular): Large multifunctional proteins; each module catalyzes one round of elongation.

Type II (Iterative): Used primarily for aromatic polyketides (*e.g.,* tetracycline).

***Type III*:** Small, standalone proteins that produce compounds like alkylresorcinols.

Key Domains: Ketosynthase (KS), Acyltransferase (AT), Ketoreductase (KR), Dehydratase (DH), Enoylreductase (ER), and Acyl Carrier Protein (ACP).

Ribosomally Synthesized and Post-translationally Modified Peptides (RiPPs)

RiPPs are synthesized on the ribosome as a precursor peptide, which is then heavily modified

Biosynthetic Logic: A precursor peptide consisting of an N-terminal leader (recognition) and C-terminal core (final product) is acted upon by enzymes that modify and cleave it.

Common Types: Lanthipeptides, lassopeptides, proteusins, thiopeptides.

Key Enzymes: Cyclodehydratases (YcaO), epimerases, and proteases.

Terpenes

Terpenes (or terpenoids) are ubiquitous natural compounds derived from 5-carbon isoprenoid units[20].

Biosynthetic Logic: Initiated by terpene synthases/cyclases that convert geranyl diphosphate (GPP), farnesyl diphosphate (FPP), or geranylgeranyl diphosphate (GGPP) into complex cyclic structures.

Precursor Pathways: Mevalonate (MVA) pathway (cytosolic) or Methylerythritol Phosphate (MEP) pathway

(plastidial).

Computational Tools Ecosystem (antiSMASH, BiG-SCAPE, MIBiG, Metabolite DBs)

This review aims to examine the bioinformatics ecosystem facilitating the integrated workflow of Anti-SMASH for biosynthetic gene cluster (BGC) identification, BiG-SCAPE for evolutionary networking, MIBiG as the comparative standard for characterized and metabolite databases with the emerging databases[16]. These computational tools transformed the search for natural products into a data-driven science[11].

antiSMASH:

The antiSMASH provides a platform for identifying Biosynthetic Gene Clusters (BGCs)[18] in the industry standard. It uses ruled-based logic to annotate core enzymes (PKS/NRPS) and tailoring genes[4]. It is essential for distinguishing common "housekeeping" metabolites from rare, niche-specific scaffolds in hotspots.

MIBiG:

The MIBiG databases provide a platform for a curated repository of experimentally validated BGCs. It serves as the "negative control" for discovery; clusters with low homology to MIBiG entries are prioritized as potential novel.

BiG-SCAPE:

It groups BGCs into Gene Cluster Families (GCFs) when dealing with hundreds of genomes from a hotspot. With BiG-SCAPE, the researchers are able to visualize the "biosynthetic landscape", which allows them to identify unique chemical families[9] that exist only in specific geographic locations or taxa.

Metabolites DBs:

The chemical evidence required for structural validation and de-replication are provided by the Metabolite Databases[21]. This integration prevents the "rediscovery trap" by cross-referencing predicted scaffolds with known chemistry.

Examples :

1. The NP Atlas (npatlas.org)
2. StreptomeDB (pharmaceutical-bioinformatics.org)
3. GNPS (gnps.ucsd.edu)
4. COCONUT[16](coconut.naturalproducts.net)

Hotspot Biogeography and Biosynthetic diversity

Research on biodiversity hotspot biogeography and biosynthetic diversity reveals that while hotspots[15] are defined by species richness and endemism, they also serve as reservoirs for unique, understudied metabolic capabilities. Key known patterns include high spatial variation in biosynthetic potential (even at local scales), strong correlation with habitat type, and intense, localized selection pressures that drive chemical diversification.

Known Patterns in Hotspot Biogeography: Hotspots cover only 2.3–2.4% of Earth's land surface yet contain ~50% of the world's endemic plant species and 43% of endemic terrestrial vertebrates. Hotspots often correspond with areas of moderate past climate variability, high elevation,and complex geology, acting as "cradles of speciation"[19] and long-term refugia. Hottest Hotspots: Analysis indicates that Madagascar, the Philippines, Sundaland, Brazil's Atlantic Forest, and the Caribbean are among the most threatened and rich, showing the highest endemism per unit area.

Latitudinal/Longitudinal Gradients: While species diversity generally peaks in tropical, low-latitude regions, specific longitudinal patterns (like the "bull's-eye" in the Indo-Australian Archipelago) are tied to tectonic history.

Biosynthetic diversity refers to the variety of natural products (specialized metabolites) a community can produce, typically encoded by Biosynthetic Gene Clusters (BGCs).Hotspots of Chemical Diversity: Soil and marine microbiomes within biodiversity hotspots are enormous, untapped reservoirs of natural products.

***Habitat-Specific Diversity*:** Biosynthetic potential strongly correlates with biome type and environmental conditions (*e.g.,* nitrogen levels in marine sediments), with algal-dominated sites showing distinct biosynthetic signatures compared to reef-dominated sites.

Chemical Novelty: Metagenomic studies (*e.g.*, in Shark Bay or Moorea) show that a large portion of detected BGCs (often >90%) are unique to specific locations and uncharacterized, representing vast "uncharted chemical space"[16].

Microbial "Dark Matter": Poorly studied phyla (*e.g.*, Acidobacteriota, Myxococcota) in these areas often encode a higher number of BGCs than known cultured strains[6], suggesting these hotspots are key for bioprospecting.

Interconnection:Biogeography+Biosynthesis and Taxonomy-Function Mismatch

Microbiome taxonomic composition does not always correlate strongly with biosynthetic potential. Finer-resolution studies are needed because closely related taxa can have vastly different metabolic capabilities[7].

Evolutionary Linkages: Hotspots that acted as long-term refugia (OCBILs—Old, Climatically-Buffered, Infertile Landscapes) are not only centers of species endemism but also potential hotspots for unique,

specialized metabolic pathways developed over long evolutionary timescales[1].

Unique Chemical Signatures: A study of marine sediments in Moorea showed that while primary metabolic genes (*e.g.*, nitrogen cycling) were redundant, specialized metabolic genes (BGCs) and the resulting compounds were highly distinct by site, driven by local interactions[16].

High-Level BGC Producers: Research in Shark Bay, Australia, revealed that Planctomycetes and Deltaproteobacteria are prolific producers of terpene and bacteriocin BGCs, which are crucial for adapting to extreme environmental conditions.

Global Mapping:Pyro-sequencing of soil microbes (nonribosomal peptide adenylation/polyketide ketosynthase domains) confirms that soil biosynthetic diversity correlates with geographic distance and biome type[1].

Future Directions (multi-omics and genomic approaches)

Multi-omics integrates genomics, transcriptomic and metabolomics to validate BGCs and to activate"cryptic" pathways. Comparative metabolomics identifies induced metabolites, reporter genes, screen activator, statistical tool like metabo Analyse dereplication known compounds.

Mining Actionmycete for Novel Antibiotics in the Omics Era-PMC

Increase in the number of partial and complete genome sequencing projects on the actinomycete species available in the public databases not only have confirmed their broad biosynthetic diversity across the different lineage but also enabled intensive genome mining approaches to untap new natural products scaffolds[1]. Relevant aspect of the impact of the increasing number of BGCs sequence information on antibiotics discovery is the possibility of developing specific targeted genome mining search in genomic libraries based on specific genomic signature related to the biosynthesis of preveleged Scaffold or functionalization that could drive the discovery of novel compounds and chemical spaces .one of the major challenges that still remain in the efficient cloning and expression of BGCs that are originally silent or poorly expressed in their natural host after the use of refactoring by the replacement of the regulatory elements and further detection of the synthesized compounds[14]. Many BGCs cannot be detected by the rules- based bioinformatic tool due to the absence of signature genes,but the application of prediction tools based on the frequencies of Pfam domains occuring in BGCs have improved the identification of additional clusters.New genomic bacterial artificial chromosomes(BAC) libraries built from large 100kb fragments of *Streptomyces* spp[13]. Genomics DNA are also used in the high throughput functional screening approaches to identify non-predicted BGCs by heterologous expression.

Genomic Approach

The discovery of *Streptomycin* from streptomyces, this genus has received considerable attention, being a primary source of antibiotics. Due to the repeated rediscovery of known compounds in the same ecological environments, as well as the associated cost[3]. Moreover, under laboratory conditions, microbes frequently cease SM production, further complicating drug discovery efforts. Many marin-derived Actinomycetota[19] genomes have been sequenced to evaluate their drug potential. Genome mining of marine sediment-derived *Streptomyces* sp. GMYO1 revealed 28 BGCs envolved in the production of flaviolin, genomic ectoine[2], class1V lanthipeptide /SFLA, albaflavenone and informatipetin. Genome mining of the deep sea- derived streptomyces antinioticus OUCT 16-23 revealed the present of fili pentypes polyene macrolids exhibiting antifungal activity against *Candida albicans*. Genome mining of marine streptomyces sp H-KF8 identified several nonribosomal peptides, leading to the design and synthesis of eight peptide,six of which showed antimicrobial[8] activity,with the two potentially disrupting membrane via a novel ion-passage mechanism. Apart from genus streptomyces, the Actinomycetota genera genome were also mined to find their secondary metabolites BGCs. Genome analysis of the genus salonispora, which was first described from a marine habitat, Nocardiopsis, harbors diverse BGCs for polypeptides, nonribosomal peptides, phenazine, bacteriocins, surfactins,and sanctipeptides[3] with many showing low similarity to known clusters, indicating potentials for novel natural products discovery.

Conclusion

This review has comprehensively examined the pivotal contributions of genome mining to revealing biosynthetic gene clusters (BGCs) for secondary metabolites in actinomycetes sourced from biodiversity hotspots, utilizing key platforms such as antiSMASH, BiG-SCAPE, MIBiG, and metabolic databases, in conjunction with multi-omics strategies and AI-powered advancements.

References

1. Adamek M, et al. Comparative genomics reveals phylogenetic distribution patterns of secondary metabolites in

Amycolatopsis species. *BMC Genomics*. 2018; **19**(1) : 426.

2. Barker DJ, Stuckey DC. A review of soluble microbial products (SMP) in wastewater treatment systems. *Water Research*. 1999; **33** : 3063–3082.

3. Belknap KC, et al. Tapping into actinobacterial genomes for natural product discovery. *Frontiers in Microbiology*. 2020; **11** : 523.

4. Blin K, et al. antiSMASH 7.0: new and improved predictions of detection, regulation, chemical structures and visualization. *Nucleic Acids Research*. 2023; **51**(W1) : W46–W50.

5. Danovaro R, et al. Major viral impact on the functioning of benthic deep-sea ecosystems. *Nature*. 2008; **454**(7208) : 1084–1087.

6. Demain AL. Importance of microbial natural products and the need to revitalize their discovery. *Journal of Industrial Microbiology & Biotechnology*. 2014; **41** : 185–201.

7. Dhanasekaran D, et al. Actinobacteria: A source of novel antibiotics from extreme environments. *Microbial Pathogenesis*. 2021; **158** : 105252. doi:10.1016/j.micpath.2021.105252.

8. Genilloud O, et al. The re-emerging role of microbial natural products in antibiotic discovery. *Antibiotics (Basel)*. 2018; **7**(4) : 84.

9. Kautsar SA, et al. BiG-SLiCE: A highly scalable tool maps the diversity of 1.2 million biosynthetic gene clusters. *GigaScience*. 2021; **10**(1) : giaa154.

10. Malik P, Rao N, Quadri SR, Sathish M, Quach NT, Li WJ, Thamchaipenet A. Exploring biosynthetic gene clusters of rare actinobacteria for novel bioactive compounds. *Frontiers in Pharmacology*. 2025; **16** :1634207. doi:10.3389/fphar.2025.1634207.

11. Medema MH, Fischbach MA. Computational approaches to natural product discovery. *Nature Chemical Biology*. 2015; **11**(9) : 639–648.

12. Moumbock AFG, et al. StreptomeDB 3.0: an updated public database of Streptomycetes-derived natural products. *Nucleic Acids Research*. 2021; **49**(D1) : D600–D605.

13. Mukherjee AK, Das K. Microbial surfactants and their potential applications: an overview. *Advances in Experimental Medicine and Biology*. 2010; **672** : 54–64.

14. Muñoz Navarro JC, et al. A computational framework to explore large-scale biosynthetic diversity. *Nature Chemical Biology*. 2020; **16**(1) : 60–68.

15. Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J. Biodiversity hotspots for conservation priorities. *Nature*. 2000; **403** : 853–858.

16. Sorokina M, Steinbeck C. COCONUT: the Collection of Open Natural Products Online database. *Journal of Cheminformatics*. 2020; **12** : 1–11.

17. Tanaka Y, Omura S. Agroactive compounds of microbial origin. *Annual Review of Microbiology*. 1993; **47** : 57–8.

18. Terlouw BR, et al. MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Research*. 2023; **51**(D1) : D603–D610.

19. Van Santen JA, et al. The Natural Products Atlas 2.0: a comprehensive resource for natural products discovery. *Nucleic Acids Research*. 2022; **50**(D1) : D1317–D1323.

20. Wang M, et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology*. 2016; **34**(8) : 828–837.

21. Weber T, Kim HU. The secondary metabolite bioinformatics portal : Computational tools to facilitate synthetic biology of secondary metabolite production. *Synthetic and Systems Biotechnology*. 2016; **1**(2) : 69–79.