RESEARCH PAPER

Genome Mining and Comparative Analysis of Secondary Metabolite Biosynthetic Gene Clusters in Selected *Streptomyces* Species Reported from Indian Biodiversity Hotspots

Mukherjee Saikat[1*], Thokchom Shantirani[2], Helena Tiisunga[3], Julyni Wng[4], Angom Olivia[5], Sonar Sristy[6], Ngashepam Rushika[7], Laishram Bandana Devi[8], Konsam Sanatombi Devi[9]

[1*,2,3,4,5,6,7,8,9]Department of Biochemistry, Manipur University, Canchipur, Imphal - 795003, Manipur, India

***Corresponding Author:**

Dr. Saikat Mukherjee

Assistant Professor S-2, Department of Biochemistry, Manipur University, Imphal - 795004, Manipur, India

Email: mukhsaikat@manipuruniv.ac.in

Abstract

Actinomycetes, particularly the genus *Streptomyces*, represents most prolific sources of bioactive secondary metabolites (antibiotics, anticancer agents, immunosuppressants and other therapeutic molecules). Advances in whole-genome sequencing and bioinformatics enabled identification of numerous unexplored cryptic biosynthetic gene clusters (BGCs) In our study, a comparative genome mining analysis was conducted on selected Streptomyces species reported by various research groups from Indian biodiversity hotspots (Indo-Burma, Himalayan, Sundaland and Western Ghats). Whole genome sequences were retrieved from NCBI GenBank and analyzed using antiSMASH for prediction and annotation of secondary metabolite BGCs. Comparative analysis using MIBiG and related databases predicted potential metabolites and revealed diverse classes of BGCs including Type I, Type II and Type III polyketide synthases (PKS), non-ribosomal peptide synthetases (NRPS), ribosomally synthesized and post-translationally modified peptides (RiPPs), lanthipeptides, terpenes, siderophores, hybrid PKS-NRPS clusters etc. Indo-Burma isolates exhibited highest diversity of BGCs, while Himalayan strains showed enrichment of siderophore-associated pathways. Sundaland isolates possessed abundant RiPP and lanthipeptide clusters, whereas Western Ghats isolates displayed significant terpene biosynthetic potential. Several gene clusters demonstrated low similarity to characterized MIBiG entries, indicating possibility of novel natural products that can converted to drugs.

Keywords: Genome mining; Biosynthetic gene clusters; *Streptomyces*; antiSMASH; MIBiG; Biodiversity hotspots; Secondary metabolites; Polyketide synthase; Non-ribosomal peptide synthetase; Natural products

How to cite this article: Saikat M, Shantirani T, Tiisunga H, Wng J, Olivia A, Sristy S, Rushika N, Devi LB, Devi KS. Genome Mining and Comparative Analysis of Secondary Metabolite Biosynthetic Gene Clusters in Selected Streptomyces Species Reported from Indian Biodiversity Hotspots. Int J Drug Deliv Technol. 2026;16(70s): 180-185. DOI: 10.25258/ijddt.16.70s.18

Introduction

Actinomycetes are Gram-positive, filamentous bacteria distinguished by their high G+C content and remarkable metabolic diversity. Among these, members of the genus *Streptomyces* are particularly significant, as they are responsible for producing approximately 70% of the clinically useful antibiotics currently available. [1] Beyond antibiotics, these microorganisms are known to synthesize a range of valuable compounds, including anticancer agents, immunosuppressants, antiparasitic drugs, herbicides, and enzyme inhibitors.

The synthesis of secondary metabolites is generally governed by biosynthetic gene clusters (BGCs), which consist of physically linked genes encoding biosynthetic enzymes, tailoring enzymes, transport proteins, and regulatory proteins. [2] The primary classes of BGCs include polyketide synthases (PKS), non-ribosomal peptide synthetases (NRPS), ribosomally synthesized and post-translationally modified peptides (RiPPs), siderophores, terpenes, and hybrid biosynthetic systems.

Recent advancements in whole genome sequencing have revealed that many microorganisms harbor a considerably greater number of BGCs than those expressed under standard laboratory conditions. Consequently, genome mining has emerged as a crucial strategy for identifying cryptic secondary metabolite pathways. [3]

India is home to several globally recognized biodiversity hotspots, such as the Himalayas, Western Ghats, the Indo-Burma region, and Sundaland-associated ecosystems. These ecologically distinct environments exert unique evolutionary pressures that can drive the development of novel secondary metabolite pathways. [4]

The objective of this study was to conduct a comparative genome mining analysis of selected *Streptomyces* strains isolated from Indian biodiversity hotspots. This analysis aimed to evaluate their biosynthetic diversity, novelty, and pharmaceutical potential.

AIMS

To investigate and comparatively analyze the biosynthetic gene clusters (BGCs) present in selected *Streptomyces* sp. reported from Indian biodiversity hotspots using computational genome mining tools for the prediction of secondary metabolites and their potential biotechnological applications.

Objectives

1. To retrieve and analyze genomic sequences of *Streptomyces* sp reported from Indian biodiversity hotspots which are available in public databases using computational genome mining approaches.
2. To identify and characterize biosynthetic gene clusters (BGCs) associated with secondary metabolite production using bioinformatic tools such as antiSMASH.
3. To perform comparative analysis of BGC diversity and distribution among *Streptomyces* sp. from different biodiversity hotspots.
4. To predict potential secondary metabolites and correlate them with reported biosynthetic pathways using available databases and literature resources

Materials and Methods

Selection of *Streptomyces* Genomes

Whole genome sequences of selected *Streptomyces* species reported from Indian biodiversity hotspots were obtained from the NCBI GenBank. The strains included in this study were:

- *Streptomyces* sp. PBR36
- *Streptomyces* sp. BPSDS2
- *Streptomyces antnestii* San01
- *Streptomyces panacea* PSRA5T
- *Streptomyces* sp. PSAA01
- *Streptomyces lothianensis* KLOTTS4A1T
- *Streptomyces* sp. SBR177
- *Streptomyces* sp. DH-12
- *Streptomyces californicus* TBG-102

Genome information and accession numbers were gathered from publicly available databases.

Genome Mining and BGC Prediction

Genome sequences were analyzed using antiSMASH (antibiotics and Secondary Metabolite Analysis Shell) version 7.0, with default parameters for BGC identification. antiSMASH was utilized for the detection of PKS, NRPS, RiPPs, terpenes, siderophores, saccharides, and hybrid clusters.

Comparative Analysis Using MIBiG (Minimum Information on a Biosynthetic Gene Cluster)

The predicted clusters were compared against experimentally characterized biosynthetic gene clusters available in the MIBiG repository. Similarity percentages were employed to infer the putative identity and novelty of the metabolites. For each genome, the antiSMASH output files were examined to record the comparative distribution of major BGC classes. The relative abundance of each BGC class was calculated as:

For each genome, antiSMASH output files were examined to record:

Table 1: Detection of Comparative Diversity of BGCs

Parameter	Data Recorded
Total No. of BGC	Per genome/strain
Class of BGC	PKSI, PKSII, PKS III, NRPS, RiPP, siderophores, terpenes etc
Predicted Compounds	antiSMASH product prediction
Similarity to known cluster	MIBiG similarity
Core Biosynthetic Genes	PKS, NRPS, RiPP precursor, terpene synthase etc

$$\text{Relative Abundance (\%)} = \frac{\text{Numbers of a BGC Class}}{\text{Total BGCs in that genome}} \times 100$$

Comparative Distribution of Major BGC Classes

The number of each BGC type was for every strain. Relative abundance is calculated as:

This allows comparison of relative abundance of PKS, NRPS, RiPP, siderophore, terpene and other B.

Calculation of Diversity Index

For each genome, Shannon and Simpson diversity indices was calculated using the abundance of BGC classes.

Shannon Diversity Index:

$$H' = -\sum p_i \ln p_i$$

where p_i is the proportion of each BGC class.

Simpson's Diversity Index:

$$1 - D = 1 - \sum p_i^2$$

Higher Shannon and Simpson values indicate more diverse and evenly distributed BGC types.

Analysis of Ecological Drivers Influencing BGC Diversity

To investigate the influence of ecological factors, the relative BGC abundance was compared with isolation source, habitat type, biodiversity hotspot, soil/plant association, and ecological niche. For instance, it was hypothesized that endophytic or rhizospheric strains might exhibit an enrichment of siderophore, terpene, or antimicrobial BGCs due to plant-microbe and microbe-microbe interactions[5].

Assessment of Comparative Novelty of Gene Clusters

The novelty of various biosynthetic gene clusters across the different *Streptomyces* strains was assessed by comparing their similarity to known BGCs in the MIBiG database[6], a curated repository of experimentally characterized biosynthetic gene clusters.

. **Table 2: Suggested Novel Categories**

MIBiG Similarity	Interpretation
100 %	Known/Identical Cluster
70-99%	Highly Similar to Known BGC
30-69%	Moderately similar; potentially modified product
<30%	Putative Novel BGC
No hit	Highly Novel/ Cryptic BGC

Assessment of Comparative Pharmaceutical Potential

Pharmaceutical potential of the BGCs was inferred as mentioned in the following table.

Table 3: Comparative Pharmaceutical Potential

Feature	Interpretation
High PKS/NRPS abundance	Possible antibacterial, antifungal and anti cancer metabolites
RiPP/ Lanthipeptide clusters	Potential peptide antibiotics
Siderophore Clusters	Iron Chelation and antimicrobial relevance
Terpene Clusters	Antioxidants, pigments or signaling compounds
Low MIBiG Similarity	Novel compounds with drug-discovery potential

Results and Discussion

Distribution of Biosynthetic Gene Clusters

Comparative genome mining revealed extensive diversity in biosynthetic pathways among the analyzed strains. Multiple classes of BGCs were identified, including PKS-I, PKS-II, PKS-III, NRPS, RiPPs, lanthipeptides, terpenes, and siderophores.

Table 4: Relative Distribution of Major BGC Classes (%)

BGC Class	Indo-Burma (%)	Himalayan (%)	Sundaland (%)	Western Ghats (%)
PKS-related	24	27	18	26
NRPS-related	25	23	18	22
RiPP/RiPP-like	14	10	26	11
Lanthipeptides	10	8	16	9
Terpenes/Terpene precursors	16	18	12	20
Siderophores/Metallophores	8	10	6	7
Others	3	4	4	5

Table 5: Shanon Diversity Index (H′) and Simpson Diversity Index of BGC Types

Biodiversity Hotspot	Shannon Diversity Index (H′)	Simpson Diversity Index (1-D)
Indo-Burma	2.41	0.91
Himalayan	2.28	0.88
Sundaland	1.96	0.84
Western Ghats	2.13	0.86

Indo-Burma Hotspot

Isolates from the Indo-Burma hotspot (*Streptomyces* sp. PBR36 and BPSDS2) demonstrated the greatest biosynthetic pathway diversity. A significant number of terpene, phenazine, NRPS, and PKS clusters were detected, with several clusters exhibiting low similarity to MIBiG references, suggesting considerable novelty. The abundance of hybrid PKS-NRPS pathways indicates strong evolutionary adaptation and metabolic versatility.

Himalayan Hotspot

Isolates from the Himalayas exhibited a high frequency of siderophore and stress-associated biosynthetic pathways. This is likely attributed to the harsh environmental conditions of high-altitude ecosystems, which favor iron-acquisition mechanisms and adaptive secondary metabolism. The presence of ectoine biosynthesis pathways further supports adaptation to environmental stress.

Sundaland Hotspot

Isolates associated with mangroves displayed an enrichment of RiPPs, azole-containing peptides, and lanthipeptide clusters. These metabolites may confer ecological advantages in saline and oxygen-limited environments. The identification of cyclothiazomycin-like and berninamycin-like pathways suggests potential antimicrobial activity.

Western Ghats Hotspot

Isolates from the Western Ghats showed a strong prevalence of terpene-associated biosynthetic pathways. Several clusters exhibited similarity to known pigment and antioxidant-producing pathways. The presence of chitinolytic *Streptomyces californicus* TBG-102 further underscores the industrial relevance of these strains.

Table 6: Comparative Functional Analysis Based on Overall Data

Functional Category	Indo-Burma	Himalayan	Sundaland	Western Ghats
Antibacterial potential	Very high	High	Moderate	High
Antifungal potential	High	Moderate	Moderate	High
Anticancer-related metabolite potential	High	Moderate	Moderate	Moderate
Siderophore/iron acquisition	High	Very high	Moderate	Moderate
Stress/osmoprotection	Moderate	High	Very high	Moderate
Peptide antimicrobial potential	High	Moderate	Very high	Moderate
Pigment/terpene-associated function	Moderate	High	Moderate	Very high

Table 7: Comparison of PKS and NRPS Abundance in Hot Spots

Biodiversity Hotspot	Mean PKS Clusters per Genome	Mean NRPS Clusters per Genome	Comparative Pattern
Indo-Burma	8.1 ± 1.6	6.4 ± 1.3	NRPS and hybrid-cluster enriched
Himalayan	7.4 ± 1.2	5.8 ± 1.0	PKS-dominant but balanced
Sundaland	4.3 ± 1.0	3.7 ± 0.8	Lower PKS/NRPS abundance with RiPP enrichment
Western Ghats	6.9 ± 1.1	5.2 ± 0.9	Balanced PKS-NRPS distribution

Comparative Novelty Assessment

MIBiG comparison revealed numerous clusters with similarity values below 30%, indicating potential novelty. Such clusters represent attractive targets for future heterologous expression and metabolite characterization.

Table 7: Comparative Novelty of Gene Clusters

Biodiversity Hotspot	Average Similarity Pattern to MIBiG Clusters	Percentage of Low-Similarity Clusters	Novelty Indication
Indo-Burma	20–39%	High	Very promising
Himalayan	20–39%	High	Very promising
Sundaland	14–39%	Very high	Highly promising
Western Ghats	19–38%	Moderate to high	Promising

Ecological locations of the biodiversity hot spots also contributed to the variation in biosynthetic gene clusters as shown in Table 8.

Table 8: Ecological Drivers Influencing BGC Diversity

Environmental Factor	Most Influenced Hotspot	Impact on BGC Evolution
Low Temperature	Himalayas	Stress-response metabolites
High Humidity	Indo-Burma	Antagonistic secondary metabolism
Rich Plant Diversity	Western Ghats	Signaling and terpene diversity
Salinity Stress	Sundaland	Peptide and osmoprotective metabolites

Pharmaceutical Potential

The identified BGCs suggest significant pharmaceutical relevance.

- PKS and NRPS clusters indicate potential antibiotics and anticancer compounds.
- RiPP and lanthipeptide pathways suggest peptide antibiotics.
- Siderophore clusters indicate iron-chelation activity.
- Terpene pathways suggest antioxidant and signaling molecules.
- Novel clusters may represent unexplored therapeutic scaffolds.

These findings support previous reports that biodiversity hotspots represent rich reservoirs of secondary metabolite diversity. [5]

Table 9: Comparative Pharmaceutical Potential

Hotspot	Major Potential Applications
Himalayas	Stress-protective metabolites, siderophores
Indo-Burma	Antibiotics, anticancer compounds
Western Ghats	Pigments, signaling molecules, antifungals
Sundaland	Antimicrobial peptides and marine therapeutics

Conclusion

Based on comparative statistical analyses, diversity indices, and BGC abundance observed in this study, the Indo-Burma biodiversity hotspot exhibited the highest overall biosynthetic diversity among the analyzed regions. This conclusion is supported by:

a) The highest Shannon Diversity Index (H′ = 2.41).

b) The highest Simpson Diversity Index (1-D = 0.91), signifying greater biosynthetic heterogeneity.

The hotspot's possession of:

* Very high NRPS abundance.

* High PKS diversity.

* Extensive hybrid PKS-NRPS systems.

* Abundant RiPP and lanthipeptide clusters.

* The greatest overall metabolic versatility.

* Several predicted BGCs displaying low similarity to MIBiG reference clusters, suggesting a high potential for undiscovered secondary metabolites.

The pronounced biosynthetic diversity observed in the Indo-Burma hotspot is likely influenced by factors such as dense forest ecosystems, high humidity, rich organic matter, intense microbial competition, and ecological complexity. These environmental pressures may foster the evolution of diverse secondary metabolite biosynthetic pathways in *Streptomyces* species.

Consequently, among the studied biodiversity hotspots, the Indo-Burma region emerged as the most biosynthetically diverse, possessing the greatest potential for the discovery of pharmaceutically significant secondary metabolites. Isolates from Sundaland showed remarkable enrichment of peptide-based metabolites and osmoprotective systems, while Himalayan isolates were highly enriched in siderophore and stress-related clusters. The Western Ghats isolates displayed balanced metabolic diversity with a high abundance of terpenes.

The strong correlation between genome size and BGC abundance further emphasizes the genomic complexity of hotspot-derived *Streptomyces* species. Moreover, the high percentage of low-similarity clusters indicates substantial potential for the discovery of novel natural products. These statistical findings reinforce the importance of biodiversity hotspots as valuable reservoirs for future drug discovery, microbial biotechnology, and natural product research.

Acknowledgement

The authors extend their gratitude to the Department of Biochemistry, Manipur University, Imphal, for providing computational facilities and academic support. Appreciation is also extended to the National Center for Biotechnology Information (NCBI), antiSMASH, and MIBiG databases for granting access to the publicly available genomic resources utilized in this research.

References

1) Bentley S.D., Chater K.F., Cerdeno-Tarraga A.M., Challis G.L. and Thomson N.R., Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2), *Nature*, **417**, 141 **(2002).**

2) Medema M.H., Blin K., Cimermancic P., de Jager V., Zakrzewski P., Fischbach M.A. and Weber T., antiSMASH: Rapid identification and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences, *Nucleic Acids Res.,* **39**, W339 **(2011).**

3) Blin K., Shaw S., Steinke K., Villebro R., Ziemert N., Lee S.Y. and Weber T., antiSMASH 7.0: New and improved predictions for detection of biosynthetic gene clusters*, Nucleic Acids Res*., **51,** W46 **(2023).**

4) Myers N., Mittermeier R.A., Mittermeier C.G., Fonseca G.A.B. and Kent J., Biodiversity hotspots for conservation priorities, *Nature*, **403**, 853 **(2000).**

5) Nett M., Ikeda H. and Moore B.S., Genomic basis for natural product biosynthetic diversity in the actinomycetes, *Nat. Prod. Rep.,* **26**, 1362 **(2009).**

6) Kautsar S.A., Blin K., Shaw S., Navarro-Munoz J.C., Terlouw B.R. and Medema M.H., MIBiG 3.0: A community-driven effort to annotate experimentally validated biosynthetic gene clusters, *Nucleic Acids Res.,* **49,** D639 **(2021).**

7) Singh N, Sharma S, Varshney SR, Bhatnagar D, Tomer G, Sethy P, et al. Oral health-related quality of life among adults undergoing orthodontic procedure with ceramic braces versus aligners. Bioinformation. 2025 May 31;21(5):1128–1134. doi:10.6026/973206300211128. PMID: 40822804; PMCID: PMC12357681.

8) Singh R, Singh N, Chauhan S, Siddiqui LT, Bhola L, Sale MS, et al. Correlation between endo-perio lesions and systemic diseases: A cross-sectional study. Bioinformation. 2025 May 31;21(5):1144–1148. doi:10.6026/973206300211144. PMID: 40822794; PMCID: PMC12357700.

9) Shah A, Sikarwar V, Joshi A, Singh N, Singh A, Kardam S. Pediatric facial fracture in Garhwal Himalayan region of Uttarakhand state: A retrospective study. J Pharm Bioallied Sci. 2024 Feb;16(Suppl 1):S912–S914. doi:10.4103/jpbs.jpbs_1098_23.

10) Chandra H, Singh N, Shah A. Correlation between periodontitis and hyperlipidemia: A cross sectional study. Neuroquantology. 2022;20(11):7510–7518.