Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Genome-Based Characterisation of Streptomyces sp. LKIT from Loktak Lake Reveals a Distinct Genomospecies with Biosynthetic Potential

Sudhansu Kumar Gouda
KIIT University
Khushbu Kumari
KIIT University
Ananta Narayan Panda
KIIT University
Dinabandhu Sahoo
University of Delhi
Vishakha Raina
vraina@kiitbiotech.ac.in

KIIT University

Research Article

Keywords: Streptomyces, Loktak Lake, Genome mining, Biosynthetic gene clusters, Genomospecies, Antimycin

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11563075/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Genome-Based Characterisation of *Streptomyces* sp. LKIT from Loktak Lake Reveals a Distinct Genomospecies with Biosynthetic Potential

Sudhansu Kumar Gouda[1], Khushbu Kumari[1], Ananta Narayan Panda[1], Dinabandhu Sahoo[2], Vishakha Raina[1]*

[1]School of Biotechnology, Kalinga Institute of Industrial Technology (KIIT) Deemed to be University, Bhubaneswar, Odisha 751024, India

[2]Department of Botany, University of Delhi, Delhi 110007, India

Corresponding author: Vishakha Raina, vraina@kiitbiotech.ac.in

ORCID: Vishakha Raina 0000-0001-6095-7494

1. Abstract

Freshwater sediment remains an underexplored source of culturable actinobacteria and their biosynthetic potential. Here we report the isolation and genome-guided characterisation of *Streptomyces* sp. LKIT (MTCC 14074), recovered from the sediment of freshwater Loktak Lake, a Ramsar wetland in Northeast India. LKIT showed the strongest antibacterial activity against Gram-positive cocci, including methicillin-resistant *Staphylococcus aureus* (MRSA) and weaker activity against tested Gram-negative bacteria. Although its 1375bp, 16S rRNA gene was 100% identical to that of *Streptomyces antibioticus* NBRC 12838^T, genome-based comparison separated LKIT from the *S. antibioticus* type strain DSM 40234^T (82.87% ANI; 26.4% dDDH). Phylogenomics supported its placement within a distinct, previously undescribed genomospecies. The 8.47 Mb draft genome comprised 67 contigs (71.6% G+C), with 99.99% completeness and 2.78% contamination. antiSMASH predicted 25 BGCs, including 12 gene-cluster families lacking characterised MIBiG equivalents and 16 absent from the DSM 40234^T genome. These findings establish LKIT as a genome-supported member of an undescribed *Streptomyces* genomospecies and highlight its biosynthetic potential in freshwater sediments.

Keywords *Streptomyces* · Loktak Lake · Genome mining · Biosynthetic gene clusters · Genomospecies · Antimycin

2. Introduction

Antimicrobial resistance (AMR) has become a major global health issue in the 21st century (Naghavi et al. 2024). Members of the ESKAPE group (*Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa,* and *Enterobacter* species), which are multidrug resistant, are listed among the priority pathogens set by the World Health Organisation (WHO), and novel therapeutic approaches are urgently required to combat them (De Oliveira et al. 2020). Over the past thirty years, antibiotic discovery pipelines have slowed considerably, while resistance mechanisms have continued to spread at pace. Together, these trends have opened a treatment gap that can only be closed through renewed efforts in natural product discovery (Hutchings et al. 2019; Lewis 2020).

The genus *Streptomyces* is a major source of clinically important antibiotics and has contributed substantially to the discovery of naturally derived antibacterial agents, including streptomycin and tetracycline (Barka et al. 2016; Chater 2016). This remarkable biosynthetic potential is associated with its large genomes (typically 6-12 Mb),

which encode diverse biosynthetic gene clusters (BGCs) responsible for the production of secondary metabolites through pathways involving polyketide synthases (PKSs), non-ribosomal peptide synthetases (NRPSs), ribosomally synthesised and post-translationally modified peptides (RiPPs), and hybrid biosynthetic systems (Rutledge and Challis 2015). Apart from their antibacterial properties, natural products derived from *Streptomyces* remain a leading source of clinically important metabolites. Only a small fraction of this potential is expressed under laboratory conditions as shown by genome sequencing; therefore, a single strain can encode dozens of clusters whose products remain unknown. Genome mining with tools such as antiSMASH, when combined with gene-cluster-family (GCF) networking, allows the biosynthetic potential of a strain to be catalogued and prioritised before any chemical work begins (Medema and Fischbach 2015; Blin et al. 2021). The products of these pathways include well-known bioactive scaffolds such as the antimycins, potent inhibitors of mitochondrial complex III (Seipke and Hutchings 2013; Guo et al. 2021; Zhang et al. 2023; Park 2025).

Genome-based studies have revealed taxonomic inconsistencies within *Streptomyces*, a genus comprising several hundred described species. The highly conserved 16S rRNA gene sequences often limit their ability to reliably distinguish closely related species (Rong and Huang 2012). Consequently, whole-genome metrics, including average nucleotide identity (ANI; approximately 95-96% species-level threshold) and digital DNA-DNA hybridisation (dDDH; 70% threshold), are widely used for genome-based species delineation (Auch et al. 2010; Meier-Kolthoff and Göker 2019). The consequences are practical as well as taxonomic. Strains that appear identical by 16S sequence can still show great variation in genomic content and metabolic capacity, and taxonomic variations have been identified in publicly available *Streptomyces* genomes (Belknap et al. 2020). Such inconsistencies complicate the identification and comparison of BGCs across evolutionary lineages. Therefore, robust genome-based taxonomic analysis is essential for accurately assessing BGC distribution and biosynthetic potential within the genus.

Despite extensive bioprospecting of soil and marine *Streptomyces*, freshwater sediments remain comparatively underexplored (Barka et al. 2016; Zothanpuia et al. 2018). Loktak Lake, in the Bishnupur district of Manipur, is the largest freshwater lake of Northeast India and a Ramsar-designated wetland. It lies within the Indo-Burma biodiversity hotspot and supports a rich diversity of aquatic flora and fauna, as well as the livelihoods of the communities around it. The lake is famous for the floating heterogeneous mass of vegetation, soil, and decaying organic matter known as phumdis, which form the world's only floating national park, Keibul Lamjao, and provide the natural habitat of the endangered Sangai deer. The sediments are enriched with decaying phumdis providing an immense load of plant biomass and nutrients in addition to those from the surrounding catchments (Sanasam et al. 2011). Steep physicochemical gradients and diverse microhabitats arise from seasonal variations in water level and regulated inflow after the construction of the Ithai barrage. In combination with the regional biogeographic isolation, these factors give rise to unique microbial communities harboring actinobacteria with unexploited biosynthetic potential. However, the culturable actinobacterial diversity of Loktak Lake sediments and their genomic potential for secondary metabolite production remain insufficiently characterised.

In this study, we report the isolation and characterisation of *Streptomyces* sp. LKIT from Loktak Lake sediment, using phenotypic observations, 16S rRNA gene sequence analysis, whole-genome relatedness metrics, and phylogenomics. We compare LKIT with closely related public genome assemblies to assess its taxonomic placement and examine its genome for functional features, including predicted BGCs, resistance-associated genes,

prophages, genomic islands, and orthologous gene clusters shared with related species. We further compare BGC families across LKIT and five related genomes to determine which elements of its predicted biosynthetic repertoire are shared within the lineage and which distinguish it from the reference strain. By integrating isolate-level antibacterial screening with genome-based taxonomy and comparative biosynthetic analysis, this work evaluates LKIT as a representative of a previously undescribed *Streptomyces* genomospecies and as a resource for future investigation of freshwater-sediment actinobacteria.

3. Materials and Methods

3.1. Sample collection and isolation of bacterial strain

Sediment samples were collected from the Ithing area of Loktak Lake (24.5593 °N, 93.8147 °E) at a water depth of approximately 13-15 feet (≈4.0-4.6 m), using a Van Veen grab sampler (KC Denmark) and hollow bamboo sticks. These samples were transported to the research laboratory where microbial analysis was carried out by homogenising the sediment samples using autoclaved mortar and pestle, and the homogenate was air dried for 1 h, establishing optimal conditions to isolate actinobacteria. Subsequently, serial dilution of the treated sediment was conducted in normal saline (0.9% NaCl), up to 10^{-5}. Portions of 0.1 mL from dilutions 10^{-2}, 10^{-3}, and 10^{-4} were spread on starch casein agar (SCA) plates and incubated at 30°C until growth was visible (Küster and Williams 1964; Williams and Davies 1965). Pure cultures were isolated and selected for further screening based on the consistency of their morphological characters upon successive subculturing. The isolates were then preserved in 20% glycerol at -20°C as stock for future use. Using a panel of Gram-positive and Gram-negative test bacteria (Table 1), the isolates underwent primary antibacterial screening using the Kirby-Bauer (Bauer et al. 1966) disc diffusion method. Each test organism, adjusted to 0.5 McFarland standard, was spread uniformly onto Luria-Bertani (LB) agar plates, and sterile 6 mm discs loaded with 10 µL of the crude extract were placed on the agar surface; ciprofloxacin (5 µg), gentamicin (10 µg) and erythromycin (15 µg) discs served as positive controls. After incubation at 37°C for 18-24 h, activity was recorded as the diameter of the zone of inhibition. Because strain LKIT demonstrated strong antibacterial activity against the test microorganisms in this screening, it was chosen for genome sequencing. The purified strain was deposited in the Microbial Type Culture Collection and Gene Bank (MTCC), CSIR-Institute of Microbial Technology, Chandigarh, India, under accession number MTCC 14074.

3.2. Molecular identification and phylogenetic analysis

Marmur's standard technique was followed to extract DNA from isolate LKIT (Marmur 1961; Wright et al. 2017). The 16S rRNA gene was amplified using universal primers 27F (5'AGAGTTTGATCATGGCTCAG 3′) and 1492R (5' TACGGCTACCTTGTTACGACTT-3′), with genomic DNA as template, and the amplicon was sequenced by the Sanger sequencing method (Lane 1991; Weisburg et al. 1991). Using the ClustalW tool, the 16S rRNA gene sequence was pairwise aligned and compared to another related *Streptomyces* sp. obtained from the EzBioCloud (Yoon et al. 2017a) database (http://eztaxone.ezbiocloud.net/). The phylogenetic tree was constructed using these aligned sequences with the help of the Neighbor-joining method (Saitou and Nei 1987) in MEGA X (Kumar et al. 2018). The evolutionary distance matrix was calculated using Kimura's two-parameter model (Kimura 1980). Bootstrap resampling method with 1000 replicates was used to evaluate the topology of the phylogenetic tree (Felsenstein 1985).

3.3. Morphological and Biochemical Characterisation

Cultural characteristics were evaluated on International Streptomyces Project (ISP) media 2-7 at 28°C for 7 days (Shirling and Gottlieb 1966). Spore chain morphology was examined by scanning electron microscopy (SEM) at CIF, OUAT, Bhubaneswar. Physiological and biochemical characterisation included growth temperature range (15-50°C), pH tolerance (4-12), NaCl tolerance (0-10%), and carbon source utilisation according to standard protocols (Williams et al. 1983). Hydrolytic enzyme production (cellulase, amylase, protease, lipase, chitinase) was tested on appropriate substrates (Cappuccino and Sherman 2005).

3.4. Genome sequencing and analysis

Genomic DNA was extracted from the bacterial culture of LKIT using the PureLink® Genomic DNA Mini Kit (Invitrogen, USA), following the manufacturer's protocol, and DNA integrity was checked by electrophoresis on a 1% (w/v) agarose gel. The quality and quantity of the DNA were determined by measuring UV absorbance at 260 nm (Agilent Cary 60 UV-Vis). DNA samples with an absorbance ratio of ~1.8 at 260 nm and 280 nm were further used for sequencing (Sambrook and Russell 2001). The whole genome sequencing of strain LKIT was performed at AgriGenome Labs Private Limited (Kochi, India). DNA paired-end library construction was performed using the NEBNext Ultra DNA Library Prep Kit following the manual provided. The whole genome sequencing of strain LKIT was performed using Illumina HiSeq 4000 paired-end technology (Bentley et al. 2008). To obtain high-quality reads, raw sequencing data were pre-processed by trimming adapter sequences and low-quality bases (average quality score < 30) with AdapterRemoval v2.3.1 (Schubert et al. 2016), removing duplicate read pairs with FastUniq v1.1 (Xu et al. 2012), and aligning reads with Bowtie2 v2.2.6 (Langmead and Salzberg 2012) to remove contaminant sequences.

3.5. Genome assembly, annotation and comparative genomics

Quality-filtered reads were assembled *de novo* with SPAdes v3.13.0 (Bankevich et al. 2012), ABySS v2.0.1 (Jackman et al. 2017), Velvet v1.2.10 (Zerbino and Birney 2008) and Unicycler v0.4.7 (Wick et al. 2017). Default k-mer sizes were used for the SPAdes and Unicycler assemblies, and a k-mer range of 31-95 for the Velvet and ABySS assemblies. The Unicycler assembly was used for all downstream analysis as it produced the best assembly statistics. Assembly quality was evaluated with QUAST v4.6 (Gurevich et al. 2013). Genome annotation used RAST (Aziz et al. 2008), Prokka (Seemann 2014) and NCBI PGAAP (Tatusova et al. 2016). Functional annotation used COG classification (eggNOG-mapper) (Cantalapiedra et al. 2021) and KEGG BlastKOALA (Kanehisa et al. 2016). ANI was calculated using the EzBioCloud ANI calculator (Yoon et al. 2017b). Digital DNA-DNA hybridisation (dDDH) values were obtained from Type Strain Genome Server (TYGS) (Meier-Kolthoff and Göker 2019) using the recommended formula. MLST phylogenomics used autoMLST (Alanjary et al. 2019) with >100 conserved core genes. Orthologous cluster comparison across six strains was performed using OrthoVenn3 (Sun et al. 2023). Genomic islands were predicted using IslandViewer4 (Bertelli et al. 2017). Prophage regions were identified using PHASTER (Arndt et al. 2016). CRISPR-Cas arrays and their associated *cas* genes were identified with CRISPRCasFinder (Couvin et al. 2018). Antibiotic resistance genes were identified using CARD-RGI (Alcock et al. 2023) under strict criteria. BGCs were identified using antiSMASH 8.0.4 (Blin et al. 2025) with strict detection mode and KnownClusterBlast against MIBiG 4.0 (Medema et al. 2015; Zdouc et al. 2025); each region was assigned its nearest characterised cluster and a KnownClusterBlast gene-level similarity score. Gene-

cluster families (GCFs) were inferred with BiG-SCAPE 2.0.3 (Navarro-Muñoz et al. 2020) in mix mode against MIBiG 4.0 (glocal alignment, cutoff 0.3, singletons included), pooling all region GenBank files from the six genomes; homology within priority clusters was assessed with clinker (Gilchrist and Chooi 2021). Assembly completeness and contamination were assessed with CheckM2 v1.1.0 (Chklovski et al. 2023). Pairwise ANI among all six genomes was additionally computed with FastANI v1.33 (Jain et al. 2018) to compare the reference deposits directly.

3.6. Nucleotide sequence accession numbers

The 16S rRNA gene sequence of *Streptomyces* sp. LKIT (1,375 bp) has been deposited in GenBank under accession number QB063908 (version QB063908.1). The whole genome shotgun project has been deposited at DDBJ/ENA/GenBank under accession number JCERQC000000000 (BioProject PRJNA1535169; BioSample SAMN63566428); the version described in this paper is JCERQC000000000.1.

4. Results

4.1. Isolation and identification

The sediment samples collected from Loktak Lake in Manipur, India, yielded a total of 20 *actinomycete* strains. The selective media and pre-treatment method were used to enrich *actinomycetes* while inhibiting the growth of rapidly proliferating bacteria. One of the isolates designated as LKIT was screened against a panel of Gram-positive and Gram-negative pathogens, namely *Escherichia coli* (MTCC 1885), *Salmonella enterica* (MTCC 733) (formerly *Salmonella typhi*), *Staphylococcus aureus* (MTCC 96), *Staphylococcus aureus* MDR/MRSA (ATCC 33592), *Enterobacter cloacae* (ATCC 13047), *Shigella flexneri* (MTCC 1457), *Bacillus cereus* (MTCC 7182), *Pseudomonas aeruginosa* (MTCC 2488), *Acinetobacter baumannii* (MTCC 9829), *Klebsiella pneumoniae* (ATCC 13883) and *Micrococcus luteus* (MTCC 106). In a repeat Kirby-Bauer disc diffusion method, strong and reproducible inhibition was confined to the Gram-positive cocci (*Micrococcus luteus, Staphylococcus aureus* (MTCC 96) and the methicillin-resistant reference strain ATCC 33592), with moderate activity against *Acinetobacter baumannii* and weak or no measurable activity against the remaining Gram-negative organisms (Fig. 1; Table 1).

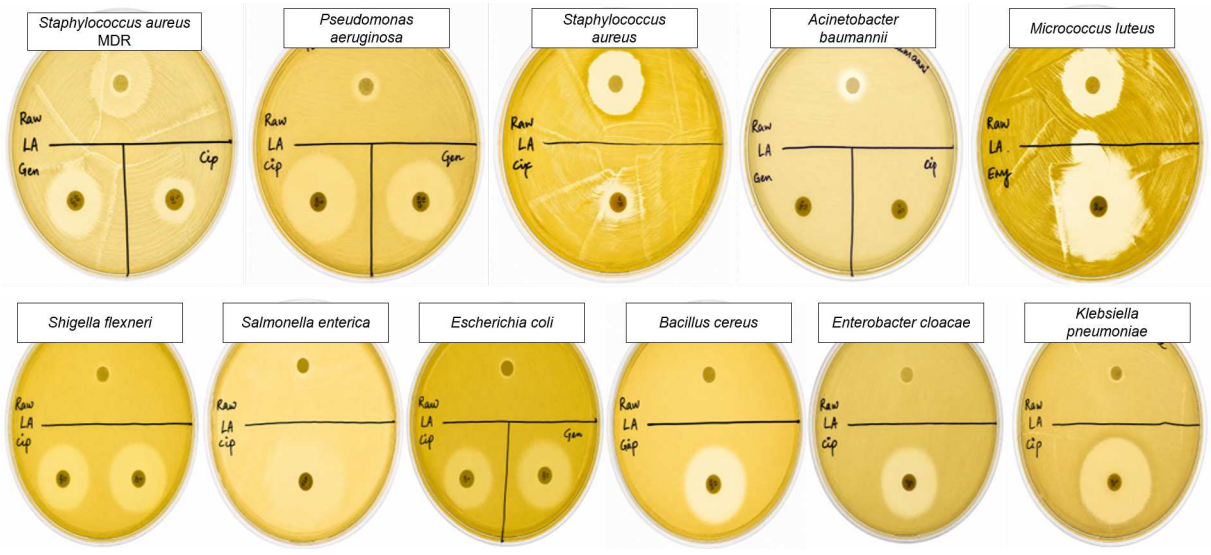


Fig. 1 Primary antibacterial screening of *Streptomyces* sp. LKIT on Luria-Bertani (LB) agar plates. Zones of inhibition produced by the LKIT crude extract (Raw) against the test panel, with reference antibiotics as positive controls (Cip, ciprofloxacin 5 µg; Gen, gentamicin 10 µg; E, erythromycin 15 µg)

Table 1. Inhibition zone profile of the *Streptomyces* sp. LKIT crude extract against the primary screening panel. Zone of inhibition diameters (mm) were measured as the total clear zone including the 6 mm disc. For Gram positive (+); Gram negative (−). A clinical antibiotic disc was included on each plate as an assay-validity control. "Interpretation" refers to susceptibility of the test organism to the LKIT crude extract.

Test organism	**Gram**	**Crude extract (zone, mm)**	**Reference antibiotic (discs)**	**Interpretation**
Staphylococcus aureus MDR/MRSA (ATCC 33592)	+	23	Gentamicin 10 µg; Ciprofloxacin 5 µg	Susceptible
Staphylococcus aureus MTCC 96	+	20	Erythromycin 15 µg	Susceptible
Micrococcus luteus MTCC 106	+	25	Erythromycin 15 µg	Susceptible
Bacillus cereus MTCC 7182	+	No zone	Gentamicin 10 µg	Not susceptible
Enterobacter cloacae ATCC 13047	−	No zone	Ciprofloxacin 5 µg	Not susceptible
Acinetobacter baumannii MTCC 9829	−	10	Gentamicin 10 µg; Ciprofloxacin 5 µg	Moderate
Pseudomonas aeruginosa MTCC 2488	−	8	Gentamicin 10 µg; Ciprofloxacin 5 µg	Marginal
Klebsiella pneumoniae ATCC 13883	−	No zone	Ciprofloxacin 5 µg	Not susceptible
Shigella flexneri MTCC 1457	−	No zone	Gentamicin 10 µg; Ciprofloxacin 5 µg	Not susceptible
Salmonella enterica MTCC 733	−	No zone	Ciprofloxacin 5 µg	Not susceptible
Escherichia coli MTCC 1885	−	Faint zone	Gentamicin 10 µg; Ciprofloxacin 5 µg	Marginal

Based on the morphological, cultural, and physiological characteristics, strain LKIT was initially assigned to the genus *Streptomyces*. The 16S rRNA gene sequence of strain LKIT (1375 bp) showed 100% sequence similarity with *Streptomyces antibioticus* NBRC 12838^{T} according to the EzTaxon database. Phylogenetic analysis performed using the neighbor-joining method in MEGA X, with 1000 bootstrap replications, revealed that strain LKIT clustered closely with NBRC 12838^{T}, supported by a bootstrap value of 100% (Fig. 2).

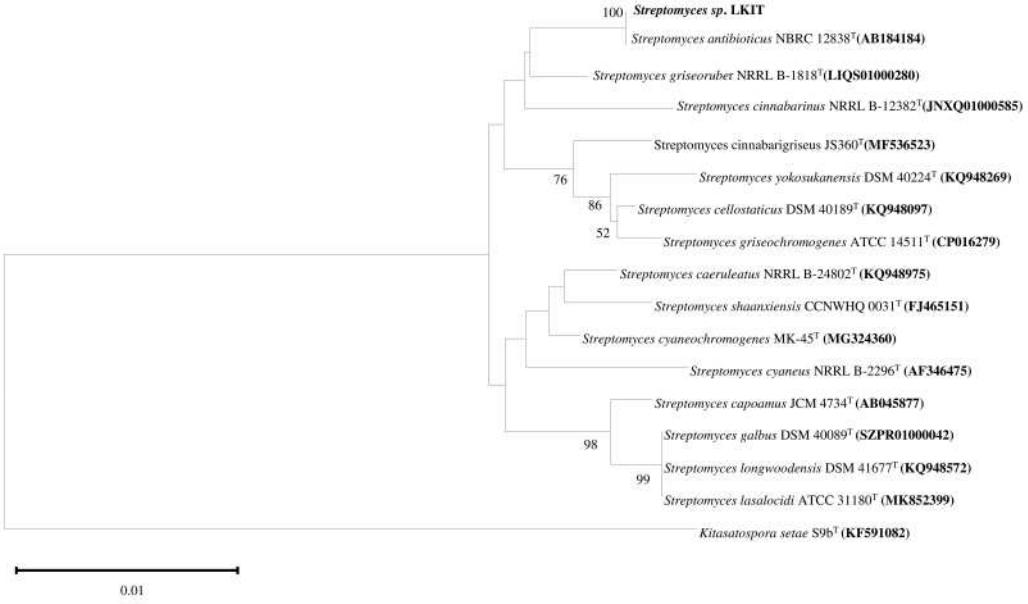


Fig. 2 Phylogenetic tree of *Streptomyces* sp. LKIT based on 16S rRNA gene sequences (1375 bp), constructed using the neighbor-joining method in MEGA X with 1000 bootstrap replicates. Bootstrap values ≥50% are shown at branch nodes. Bar = 0.01 substitutions per nucleotide position

4.2. Morphological and Biochemical Characteristics

Strain LKIT exhibited distinct growth characteristics on ISP media, with good growth and well-developed aerial mycelium on ISP3 and ISP7, and moderate growth with sporulation on ISP2, ISP4, and ISP5. No diffusible pigments were produced on ISP media, though a characteristic yellow pigment was produced on starch casein agar. A summary of cultural characteristics is presented in Table 2**.**

Table 2. Cultural characteristics of *Streptomyces sp.* LKIT on ISP media.

Medium	Growth	Substrate mycelium	Colony colour	Soluble pigment
ISP2	Moderate, sporulation	Brown	Brown	None
ISP3	Good, aerial mycelium	Dark brown	Greenish brown	None
ISP4	Moderate, sporulation	Black	Purple	None
ISP5	Moderate, sporulation	Yellowish grey	Greyish white	None
ISP6	Very limited	White	White	None
ISP7	Good, aerial mycelium	Green	Grey	None

Scanning electron microscopy showed uniform spores arranged on aerial hyphae with short, straight to flexuous spore chains, which is consistent with rectiflexible-type morphology, typical of the genus *Streptomyces* (Fig. 3). Biochemical characterisation confirmed growth from 25 to 45°C (optimum 28°C), pH 5-10 (optimum pH 7), and utilisation of a broad range of carbon sources including maltose, sucrose, ribose, dextrose, cellobiose, raffinose, xylose, fructose, arabinose, and inositol. The strain degraded chitin, cellulose, starch, casein, lipids, and gelatin, and gave positive results for catalase, citrate utilisation, methyl red, and oxidase.

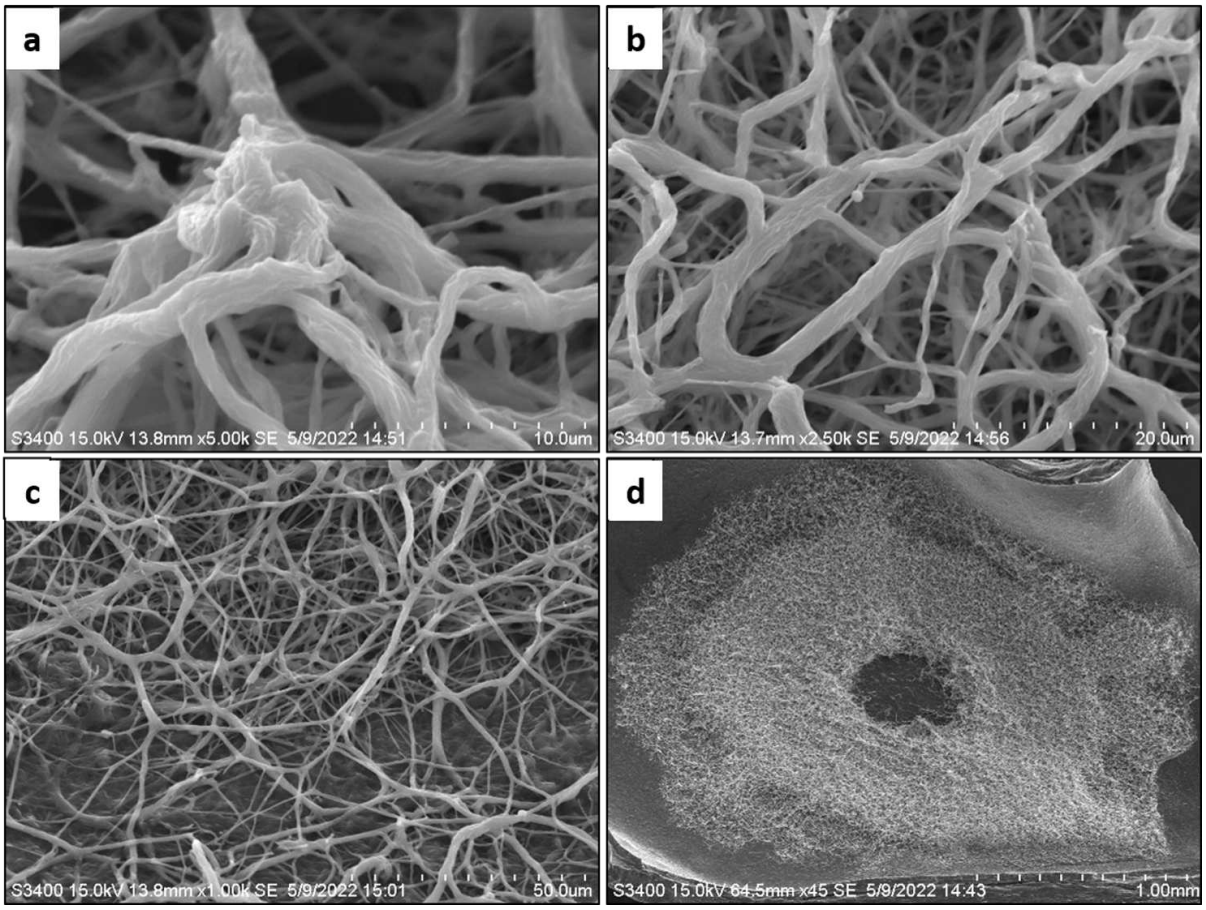


Fig. 3 Scanning electron microscopy (SEM) images of *Streptomyces* sp. LKIT showing filamentous hyphal morphology at increasing fields of view. (a) Densely branched, twisted ribbon-like hyphae at ×5.00k magnification (scale bar = 10 µm). (b) Interwoven hyphal network at ×2.50k magnification (scale bar = 20 µm). (c) Extended mycelial mat at ×1.00k magnification (scale bar = 50 µm). (d) Low-magnification overview of the mycelial mass at ×45 magnification (scale bar = 1 mm), showing overall surface topology. Images were captured at 15.0 kV acceleration voltage using secondary electron (SE) detection (Hitachi S-3400N)

4.3. Whole genome analysis of *Streptomyces* sp. LKIT

Table 3. Genome features of *Streptomyces* sp. LKIT.

Feature	**Value**
Total contigs	67
Total genome size	8.47 Mb (8,469,611 bp)
G+C content (%)	71.6

N50 (bp)	327,981
Completeness / contamination (CheckM2)	99.99% / 2.78%
Protein-coding sequences (RAST)	7,873
rRNA / tRNA / tmRNA	4 / 98 / 1
Biosynthetic gene clusters (antiSMASH 8.0.4)	25
Sequencing platform	Illumina HiSeq 4000, paired-end
Genome coverage	~1018×

The genome assembly of *Streptomyces* sp. LKIT comprises 67 contigs totaling 8,469,611 bp (8.47 Mb; Table 3), with an N50 of 327,981 bp and a G+C content of 71.6%, with an estimated genome coverage of approximately 1,018× (Table 3; Online Resource 1, Table S1). CheckM2 estimated 99.99% completeness and 2.78% contamination, indicating a highly complete draft. RAST annotation predicted 7,873 protein-coding sequences, 4 rRNA genes, 98 tRNA genes and 1 tmRNA, and the circular genome map is shown in Fig. 4. Phylogenomic analysis with autoMLST placed LKIT in a well-separated clade with *Streptomyces* sp. Amel2xC10 and *S. antibioticus* DSM 41481, distinct from the *S. antibioticus* type strain DSM 40234^T (Fig. 5). The closely related genomes NPDCO48228 and NRRL B-1701 were not included in the autoMLST reference set. They were added to the subsequent ANI, dDDH, OrthoVenn3 and BiG-SCAPE analyses because their ANI and dDDH values with LKIT were highest (Section 4.4).

4.4. Whole genome taxonomy: ANI, dDDH and phylogenomics

LKIT shared its highest average nucleotide identity with the undescribed genome *Streptomyces* sp. NPDCO48228 (98.86%; dDDH 90.6%), and exceeded the 95-96% species threshold with DSM 41481 (96.28%) and NRRL B-1701 (96.19%) whereas Amel2xC10 (95.67%) lies at the species boundary, together forming a coherent genomic cluster. Identity with the 16S matched type strain *S. antibioticus* DSM 40234^T, in contrast, was only 82.87% (dDDH 26.4%), placing LKIT outside *S. antibioticus* despite identical 16S rRNA sequences. No validly described type strain reached the 70% dDDH boundary, the closest being *S. phaeofaciens* JCM 4125^T at 35.7%. These relationships are summarised in Table 4, and the genome-based phylogeny in Fig. 5.

Table 4. Whole genome relatedness to reference genomes (OrthoANI; dDDH computed with the TYGS-recommended formula (d4)).

Reference genome	**ANI (%)**	**dDDH (%)**	**Relationship**
Streptomyces sp. NPDCO48228	98.86	90.6	Same genomic species (closest)
S. antibioticus DSM 41481	96.28	77.0	Same species-level cluster
Streptomyces sp. NRRL B-1701	96.19	78.2	Same species-level cluster
Streptomyces sp. Amel2xC10	95.67	75.7	Same species-level cluster
S. antibioticus DSM 40234^T	82.87	26.4	Different species (16S-matched type)
S. phaeofaciens JCM 4125^T	88.47	35.7	Closest type strain (TYGS)

The dDDH values from TYGS (recommended d4 formula) confirmed that all comparisons with validly named type strains fell well below the 70% species-delineation threshold, ranging from 26.4% for *S. antibioticus* DSM 40234^T (95% CI 24.0–28.9%) to 35.7% for the closest type strain, *S. phaeofaciens* JCM 4125^T. In contrast, the

values for NPDCO48228, NRRL B-1701, DSM 41481 and Amel2xC10 (75.7–90.6%) exceeded the threshold. This discordance between 16S rRNA identity and whole genome relatedness exemplifies the well-documented limits of single-gene phylogeny for species delineation in *Streptomyces* (Rong and Huang 2012), and the ANI and dDDH data place LKIT within a distinct genomospecies (Konstantinidis and Tiedje 2005).

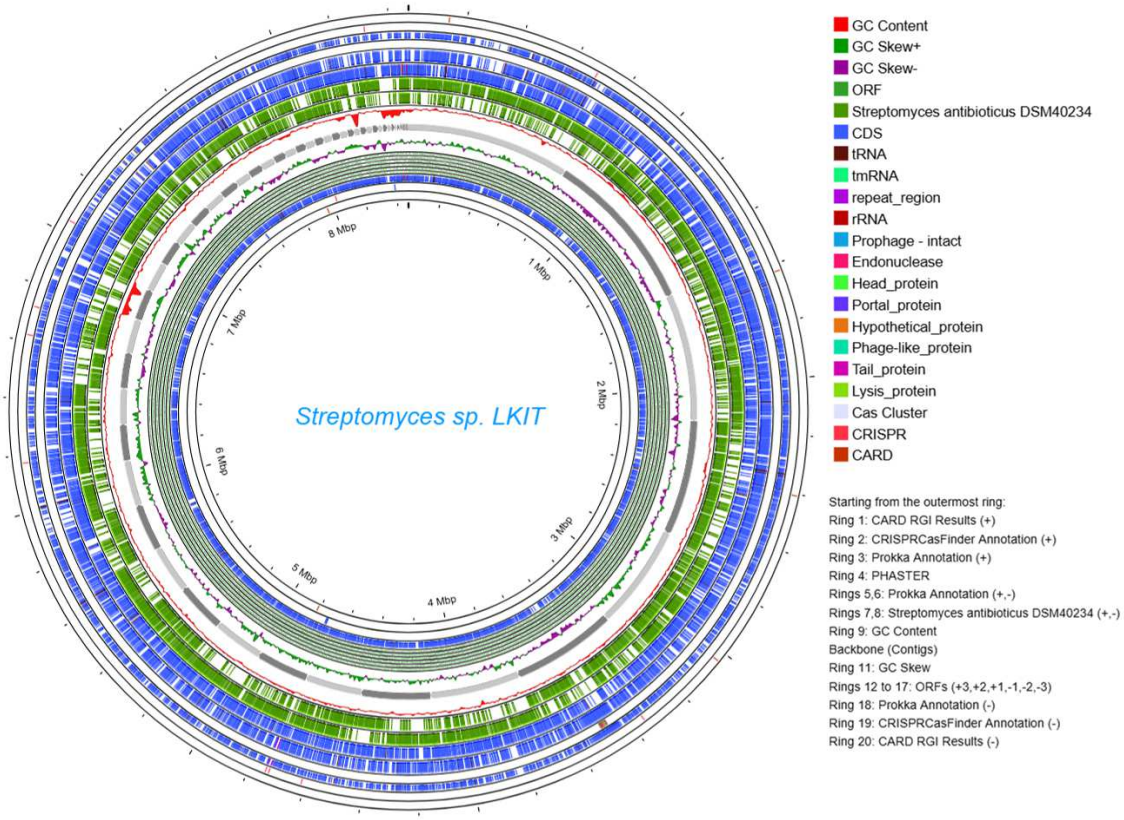


Fig. 4 Circular genome map of *Streptomyces* sp. LKIT generated by CGView/Proksee, showing GC content, GC skew, and annotated coding sequences on forward and reverse strands. Rings from outside: CARD-RGI hits, CRISPRCasFinder, Prokka annotation, PHASTER prophage and BLAST comparison with *S. antibioticus* DSM 40234^{T}

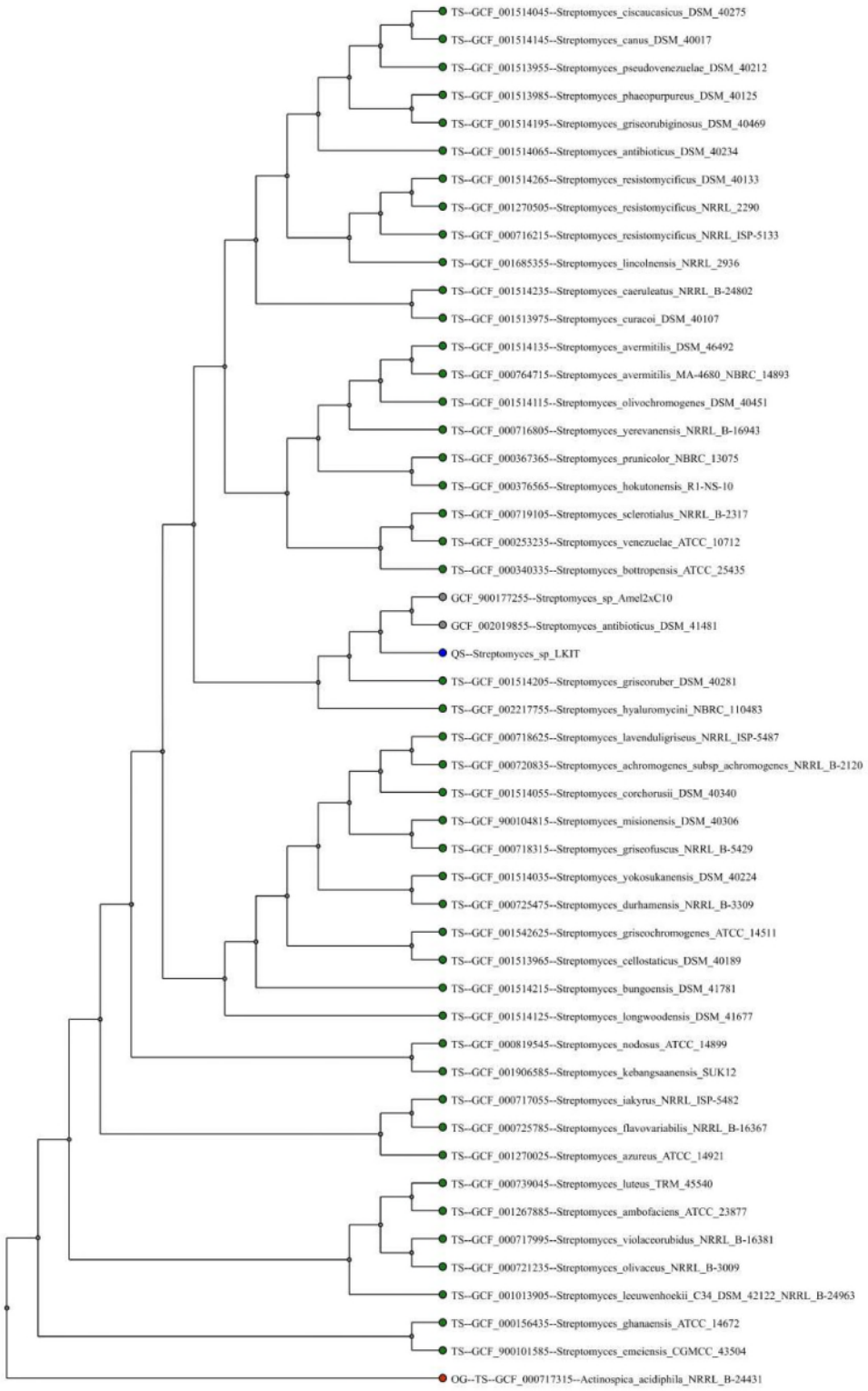


Fig. 5 Phylogenomic tree of *Streptomyces* sp. LKIT constructed with autoMLST from >100 conserved core genes; LKIT in blue, outgroup in red

4.5. Functional Genome Annotation

Of the predicted proteins, 7,087 were assigned to Clusters of Orthologous Groups (COG) categories (Fig. 6a) and 7,823 genes to KEGG pathways (Fig. 6b).

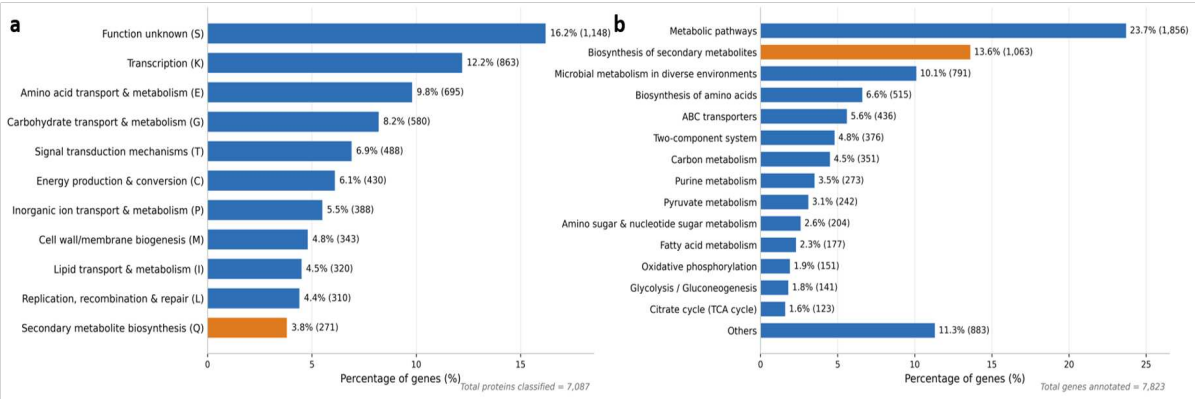


Fig. 6 Functional genome annotation of *Streptomyces* sp. LKIT. (a) COG functional classification (7,087 proteins classified). (b) KEGG pathway classification (7,823 genes annotated). The secondary metabolite category is highlighted in orange in each panel

The largest COG category was "function unknown" (S; 16.2%), followed by transcription (K; 12.2%), amino acid transport and metabolism (E; 9.8%), carbohydrate transport and metabolism (G; 8.2%), signal transduction mechanisms (T; 6.9%), and energy production and conversion (C; 6.1%). A total of 271 genes (3.8%) were assigned to COG category Q (secondary metabolite biosynthesis, transport, and catabolism).

KEGG reconstruction recovered complete central metabolic pathways (Fig. 6b), including glycolysis, the tricarboxylic acid (TCA) cycle, oxidative phosphorylation, amino acid biosynthesis, fatty acid metabolism, and carbon metabolism. Precursor-supply genes for polyketide and non-ribosomal peptide biosynthesis were present, including the fatty acid synthase genes *fabD*, *fabF*, *fabG*, and *fabH*, acyl carrier protein (*acpP*), and malonyl-CoA synthetase (*matB*). The desferrioxamine siderophore biosynthetic genes *desA*, *desB*, *desC*, and *desD* were also identified.

The genome encoded 508 transcriptional regulators, 488 signal transduction-associated proteins, approximately 280 transporter-associated proteins (including ABC transporters and efflux systems), and 14 cytochrome P450 proteins.

4.6. Comparative Genomics

4.6.1. Prophage Regions, Genomic Islands, and CRISPR-Cas

PHASTER identified three prophage regions: one intact region (53,187 bp, 64 genes, most similar to *Streptomyces* phage TG1, score 150), one incomplete region (21,427 bp, 28 genes, *Streptomyces* phage phiCAM, score 70), and one questionable region (31,654 bp, 42 genes, *Streptomyces* phage VWB, score 90). The intact prophage harbored phage integrases, tail proteins, and capsid proteins. Together, the three prophage regions span 106,268 bp (~1.3% of the 8.47 Mb genome), below the 2-8% typically reported for *Streptomyces*, suggesting a genome with limited recent prophage integration.

IslandViewer4 predicted 27 genomic islands distributed throughout the LKIT genome (Online Resource 1: Table S4), collectively encoding 456 open reading frames. These islands were enriched in mobile genetic element-associated genes, including transposases and phage-related proteins, alongside transport and metabolic functions such as ABC transporters, siderophore/iron-uptake systems, and efflux pumps. The largest islands (~324 kb and ~31 kb) were particularly rich in mobile element-associated genes, consistent with horizontally acquired genomic regions (Langille et al. 2010). CRISPRCasFinder detected 12 candidate CRISPR arrays across 10 contigs, but no cas cluster was identified.

4.6.2. Antibiotic Resistance Genes

CARD-RGI analysis under strict criteria identified six antibiotic resistance determinants in the LKIT genome (Table 5; Online Resource 1: Table S5): three glycopeptide resistance genes (*vanW*, *vanH*, *vanT*), two rifamycin resistance genes (*rox*, *helR*), and one tetracycline resistance gene (*otr(A)*). The *vanW*, *vanH*, and *vanT* genes constitute a partial *van*-type glycopeptide resistance locus and were all low-identity hits (29-38% amino-acid identity). The *rox* gene, encoding a rifampin monooxygenase, and *helR* were higher-identity hits (78.74% and 77.51% identity, respectively), and *otr(A)*, encoding a ribosomal protection protein, matched at 51.77% identity.

Table 5. Antibiotic resistance genes identified in *Streptomyces* sp. LKIT by CARD-RGI (strict criteria). The three van hits are low-identity (29-38%) intrinsic homologs; coverage >100% for vanW reflects a partial/over-length alignment.

Gene	**AMR family**	**Drug class**	**Resistance mechanism**	**% Identity**	**% Coverage**
vanW (vanI cluster)	Glycopeptide resistance cluster	Glycopeptide antibiotic	Target alteration	29.18	153.89
vanH (vanF cluster)	Glycopeptide resistance cluster	Glycopeptide antibiotic	Target alteration	37.54	102.8
vanT (vanG cluster)	Glycopeptide resistance cluster	Glycopeptide antibiotic	Target alteration	30.77	60.25
rox (*S. venezuelae*)	Rifampin monooxygenase	Rifamycin antibiotic	Antibiotic inactivation	78.74	100.21
HelR	Helicase-like RNA pol. protection protein	Rifamycin antibiotic	Target protection	77.51	99.31
otr(A) (*S. rimosus*)	Tetracycline ribosomal protection	Tetracycline antibiotic	Target protection	51.77	98.34

4.6.3. Orthologous clusters and pan-genome

OrthoVenn3 comparison of the six genomes grouped 45,392 proteins into 7,942 clusters, of which 4,741 were shared by all six (core) and 4,551 were singletons (10.03%) (Fig. 7). LKIT contained 6,943 clusters, 858 singleton proteins, and nine species-specific clusters (Table 6). When the two closest genomes (NPDCO48228, NRRL B-1701) were included, LKIT's species-specific cluster and singleton counts were lower than in comparisons lacking them. Functional annotation of the nine LKIT specific clusters identified transcriptional regulators (clusters 6769, 6770), dipeptide (7310) and ion (7315) transport, alkanesulfonate monooxygenase activity (sulfur metabolism;

7312), and a putative permease (7313); three clusters were unannotated (6768, 7311, 7314) (Online Resource 1: Table S9). None of the nine encoded a biosynthetic function, and no biosynthetic gene cluster was unique to LKIT.

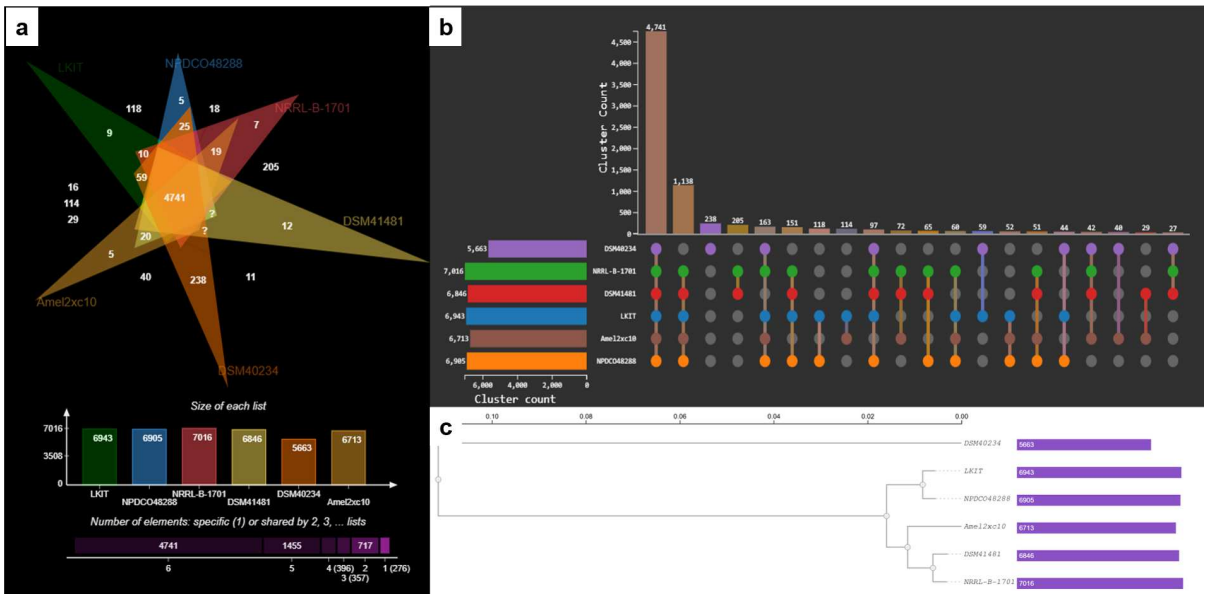


Fig. 7 Orthologous gene cluster comparison of *Streptomyces* sp. LKIT and five related genomes (NPDCO48228, NRRL B-1701, DSM 41481, DSM 40234^{T} and Amel2xC10) using OrthoVenn3. (a) Flower plot showing shared and strain-specific orthologous gene clusters. (b) UpSet plot illustrating the sizes of shared and unique cluster intersections. (c) Phylogenetic tree showing the genomic relationships among the six strains, with corresponding cluster counts

Table 6. Orthologous cluster statistics of *Streptomyces* sp. LKIT and closely related genomes generated using OrthoVenn3.

Species	Proteins	Clusters	Singletons	Species-specific clusters
Streptomyces sp. LKIT	7,873	6,943	858	9
Streptomyces sp. NPDCO48228	7,153	6,905	187	5
Streptomyces sp. NRRL B-1701	7,212	7,016	144	7
S. antibioticus DSM 41481	7,207	6,846	287	12
Streptomyces sp. Amel2xC10	7,166	6,713	381	5
S. antibioticus DSM 40234^{T}	8,781	5,663	2,694	238

4.7. Biosynthetic Gene Cluster Analysis

antiSMASH 8.0.4 (strict) identified 25 BGCs across 15 classes (Table 7; Online Resource 1: Table S2). The single strongest characterised prediction was Region 6.2 (NRPS + T1PKS, 52.6 kb), matching the antimycin A1B cluster (BGC0001455) at 100% gene-level similarity. Manual annotation resolved a complete antimycin biosynthetic logic: a threonine-loading NRPS module (adenylation-domain Stachelhaus signature DFWNVGMVHK), a malonyl-CoA PKS module terminated by a thioesterase, an aromatic 3-formamidosalicylate head from the tryptophan/kynurenine route, and a crotonyl-CoA carboxylase/reductase supplying the variable acyl side chain (Fig. 8; Fig. 9; Online Resource 1: Table S6). Further characterised predictions included a fredericamycin-type

T2PKS (Region 3.1, 81%), an ε-poly-L-lysine cluster (Region 11.1, 100%) (Yoshida and Nagasawa 2003), a γ-butyrolactone/A-factor autoregulator (Region 8.1, 100%) and an ectoine operon (Region 16.1, 100%). Adenylation- and acyltransferase-domain substrate predictions for all NRPS- and PKS-containing regions are summarised in Online Resource 1 (Table S7), providing monomer-level predictions consistent with the manual annotation of Region 6.2.

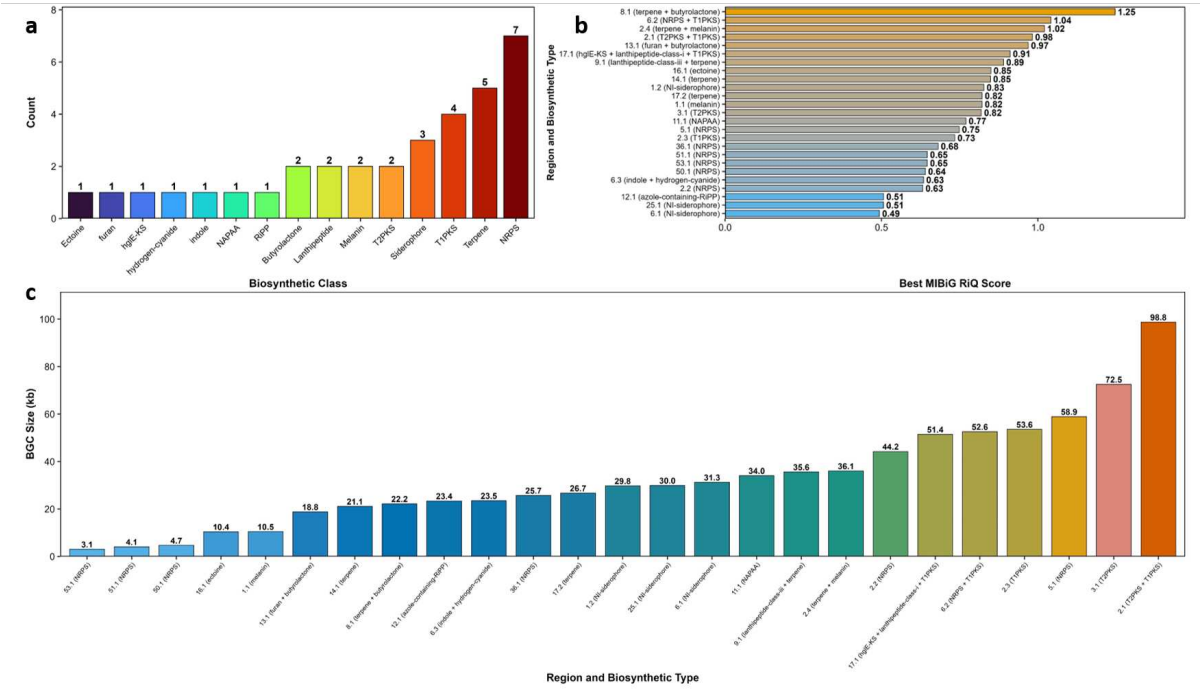


Fig. 8 Graphical representation of the biosynthetic gene clusters (BGCs) identified in the genome, illustrating the diversity, abundance, similarity scores, and size distribution of different biosynthetic classes: (a) BGC class distribution. (b) Best MIBiG cluster-comparison (RiQ) score per region from antiSMASH ClusterCompare. (c) Cluster size (kb) of all 25 BGCs

Table 7. Biosynthetic gene clusters predicted by antiSMASH 8.0.4 (strict). Full per-region detail in Online Resource 1 (Table S2).

Region	Cluster type	kb	Best MIBiG	Nearest cluster	Similarity
1.1	melanin	10.5	BGC0000700.5	*istamycin*	5%
1.2	NI-siderophore	29.8	BGC0002690.2	*FW0622*	62%
2.1	T2PKS; T1PKS	98.8	BGC0000271.5	*spore pigment*	83%
2.2	NRPS	44.3	BGC0001574.5	*SF2768*	50%
2.3	T1PKS	53.6	BGC0000873.3	*A-201A*	6%
2.4	terpene; melanin	36.1	BGC0000908.3	*melanin*	71%
3.1	T2PKS	72.5	BGC0000224.4	*fredericamycin A*	81%
5.1	NRPS	58.9	BGC0000423.5	*scabichelin*	80%
6.1	NI-siderophore	31.3	BGC0001732.5	*paulomycin*	11%
6.2	NRPS; T1PKS	52.6	BGC0001455.6	*antimycin A1B*	100%

6.3	indole; hydrogen-cyanide	23.5	BGC0002805.2	*Azodyrecin A/Azodyrecin B/Azodyrecin C*	17%
8.1	terpene; butyrolactone	22.2	BGC0000850.5	*γ-butyrolactone*	100%
9.1	lanthipeptide-class-iii; terpene	35.6	BGC0001456.3	*isorenieratene*	100%
11.1	NAPAA	33.9	BGC0002174.2	*ε-Poly-L-lysine*	100%
12.1	azole-containing-RiPP	23.4	—	—	—
13.1	furan; butyrolactone	18.8	BGC0000914.5	*methylenomycin A*	14%
14.1	terpene	21.1	BGC0000660.5	*albaflavenone*	100%
16.1	ectoine	10.4	BGC0000853.3	*ectoine*	100%
17.1	hglE-KS; lanthipeptide-class-i; T1PKS	51.4	BGC0002497.3	*hexacosalactone A*	13%
17.2	terpene	26.7	BGC0000663.5	*hopene*	92%
25.1	NI-siderophore	30	BGC0000236.5	*kinamycin*	22%
36.1	NRPS	25.6	BGC0000296.5	*actinomycin D*	28%
50.1	NRPS	4.7	BGC0002638.3	*misaugamycin A/misaugamycin B*	13%
51.1	NRPS	4.1	—	—	—
53.1	NRPS	3.1	—	—	—

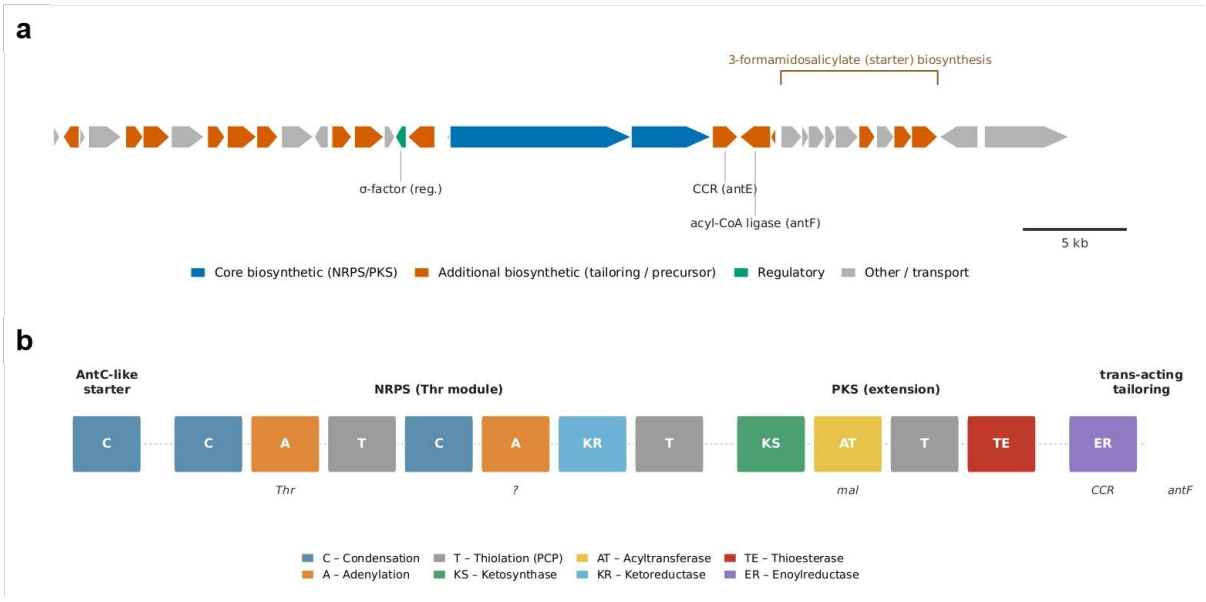


Fig. 9 (a) Biosynthetic gene cluster of Region 6.2 (antimycin-type; MIBiG BGC0001455, 100% similarity). (b) Domain architecture of NRPS-PKS assembly line

4.8. Comparative biosynthetic analysis (BiG-SCAPE)

Pooling 148 BGCs from the six genomes, BiG-SCAPE resolved 53 GCFs, 25 of which contained an LKIT cluster (Fig. 10; Online Resource 1: Table S3). Twenty-two were shared with at least one close relative; the three LKIT-only families corresponded to the contig-edge-truncated NRPS fragments (Regions 50.1, 51.1, 53.1). Two hybrid

clusters, the hglE-KS–T1PKS–lanthipeptide-class-I hybrid (Region 17.1; FAM_00144) was shared across the genomospecies, whereas the azole-containing RiPP (Region 12.1; FAM_00129) was shared only with NPDCO48228; neither family contained a MIBiG member; Clinker identified 25-31 and 14-21 homologous genes, respectively, with relatives (Online Resource 1: Table S8). Excluding the three contig-edge fragments, 12 of the 25 LKIT-containing GCFs had no match to any experimentally characterised cluster in the MIBiG database, and 16 of the 25 were absent from authentic *S. antibioticus* DSM40234^{T}. The relationships are shown as a similarity network in Online Resource 1 (Fig. S2).

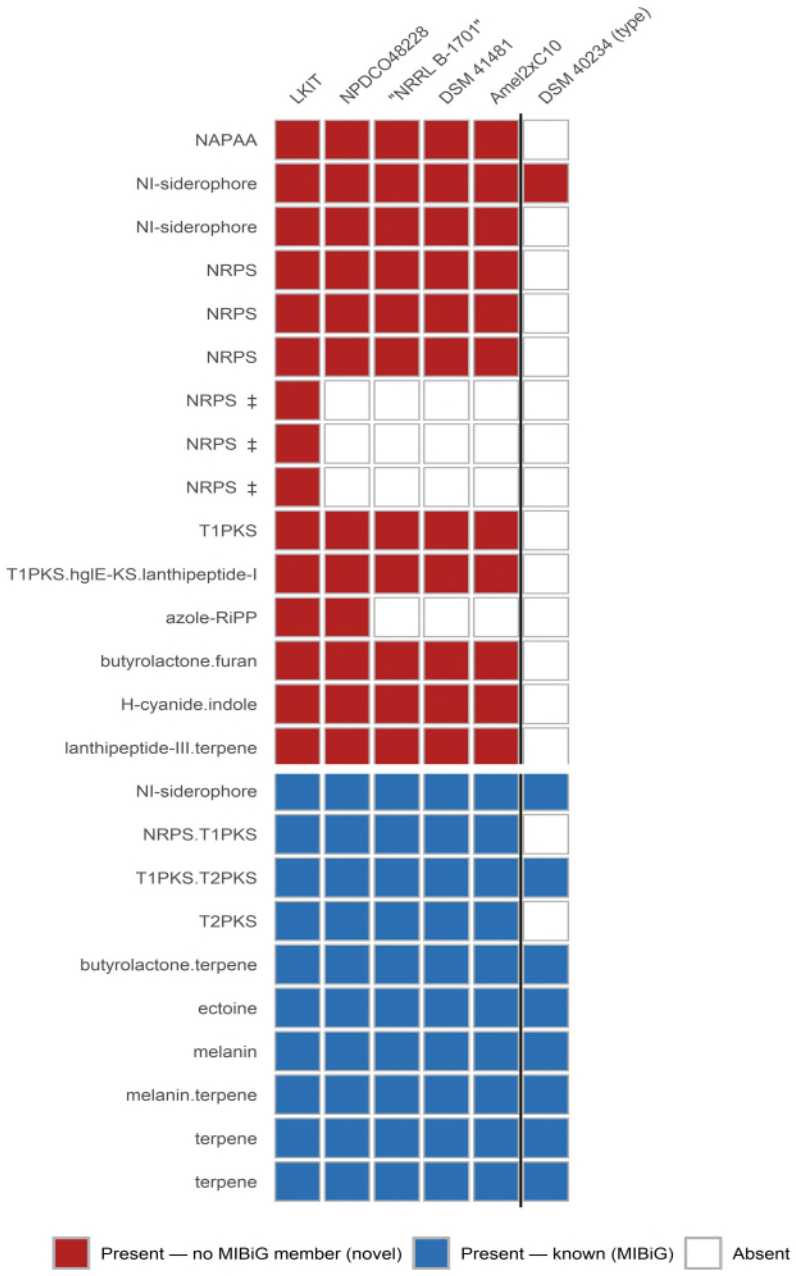


Fig. 10 BiG-SCAPE gene-cluster families containing LKIT BGCs (cutoff 0.3). Red = family with no MIBiG member (uncharacterised); blue = known family; white = absent. Columns 1-5 = the undescribed genomospecies; column 6 = authentic *S. antibioticus* DSM40234^{T} and ‡ = contig-edge truncated NRPS fragment

5. Discussion

5.1. Genome-based delineation of a previously undescribed *Streptomyces* genomospecies

The principal finding of this study is the identification of a species-level genomic lineage within *Streptomyces* that is not assignable to any of the validly named species represented in the present comparison. Whole genome analyses demonstrate that strain LKIT cannot be assigned to *Streptomyces antibioticus*, despite complete 16S rRNA gene sequence identity with the type strain *S. antibioticus* NBRC 12838^T over the 1,375-bp region analysed. This inconsistency indicates the limited discriminatory power of the 16S rRNA gene as a species-level marker in *Streptomyces*, a genus in which closely related but genomically distinct taxa may share identical or near-identical 16S rRNA gene sequences (Rong and Huang 2012). Multilocus and genome-scale approaches are therefore needed for reliable species delineation of *Streptomyces*.

LKIT showed only 82.87% ANI and 26.4% dDDH (TYGS formula d4) relative to the authentic type-strain genome of *S. antibioticus* DSM 40234^T. These values are substantially below the generally accepted genomic thresholds for conspecificity, i.e. approximately 95-96% ANI and 70% dDDH (Richter and Rosselló-Móra 2009; Chun et al. 2018; Meier-Kolthoff and Göker 2019). Thus, LKIT is genomically distinct from *S. antibioticus* DSM 40234^T and cannot be regarded as a member of that species. The result is further strengthened by comparisons with validly named type strains. The closest type strain recovered by TYGS, *Streptomyces phaeofaciens* JCM 4125^T, showed a dDDH value of only 35.7% with LKIT, again markedly below the 70% species boundary. Therefore, no currently represented validly named species provides a genomic match for LKIT.

Conversely, LKIT formed a distinct species-level genomic cluster with several publicly available genomes: *Streptomyces* sp. NPDCO48228, *S. antibioticus* DSM 41481, *Streptomyces* sp. NRRL B-1701, and *Streptomyces* sp. Amel2xC10. The strongest genomic similarity was observed with the undescribed strain NPDCO48228, with 98.86% ANI and 90.6% dDDH. LKIT also exceeded species-level ANI and dDDH thresholds with DSM 41481 (96.28% ANI; 77.0% dDDH), and NRRL B-1701 (96.19% ANI; 78.2% dDDH), whereas Amel2xC10 was at the ANI boundary and above the dDDH threshold (95.67% ANI; 75.7% dDDH). These values support the interpretation that LKIT and these strains represent a single, coherent genomic species that is distinct from all validly described *Streptomyces* species currently included in the comparison (Konstantinidis and Tiedje 2005; Chun et al. 2018).

Accordingly, LKIT should be designated *Streptomyces* sp. LKIT, representing a member of a novel genomospecies. The term “genomospecies” seems appropriate at this stage because genome-scale evidence clearly establishes species-level coherence within the LKIT–NPDCO48228 lineage and separation from named taxa, whereas formal valid publication of a new species name requires a polyphasic taxonomic description.

5.2. Inconsistent reference genome assignments and their taxonomic implications

The public genome records labelled as *S. antibioticus* require cautious interpretation. Although NRRL B-1701 and JCM 4620 are catalogue-listed equivalents of the *S. antibioticus* type strain DSM 40234^T, their deposited genome assemblies do not show the genomic relatedness expected among authentic subcultures of a single type strain. Culture-collection catalogues list DSM 40234^T, JCM 4620^T, NRRL B-1701^T, and NBRC 12838^T as strain equivalents of the *S. antibioticus* type strain (Parte et al. 2020). However, FastANI comparison showed that the

genome deposited as *S. antibioticus* NRRL B-1701 (GCA_053548315) shared only 85.4% ANI with the authenticated DSM 40234^T type strain genome (GCA_001514065), a value that is incompatible with membership in the same species (Online Resource 1: Table S10, Fig. S1) (Jain et al. 2018). In contrast, NRRL B-1701 displayed 96.9% ANI to LKIT and clustered with LKIT, NPDCO48228, DSM 41481, and Amel2xC10. The genomes assigned to NPDCO48228 and DSM 41481 showed the same pattern, i.e. approximately 85% ANI to DSM 40234^T but approximately 97-99% ANI to LKIT and its close relatives.

Independent genome sequences derived from the same type strain would normally be expected to show near-complete genomic similarity. The approximately 85% ANI observed between the deposited NRRL B-1701 and DSM 40234^T assemblies indicates a likely inconsistency in strain identity, genome labelling, or sequence source. The discordance between culture-collection designation and genome-scale relatedness indicates that these genome assemblies should be treated as putatively misidentified, inconsistently labelled, or of uncertain source pending resequencing of authenticated culture material. For taxonomic comparison, DSMZ-curated and TYGS-validated DSM 40234^T was retained as the authoritative reference genome for *S. antibioticus* because it is clearly deposited as the species type strain, its genome project identifies it as the type strain, and it is consistent with the type-strain designation maintained by DSMZ and other major culture-collection catalogues. Consequently, the placement of LKIT outside *S. antibioticus* is based on comparison with DSM 40234^T, rather than on genome assemblies whose identity or source is contradicted by their whole-genome relationships. The taxonomic significance of LKIT is based on two complementary observations: genomic coherence with several close relatives and clear distinction from the validly named species examined. It shows that a close relationship to previously deposited genomes does not rule out taxonomic novelty; instead, it suggests that the lineage has been identified but remains poorly characterised and taxonomically unresolved. Based on available evidence, LKIT can be described as a potential novel *Streptomyces* genomospecies. Formal description of the LKIT genomospecies will require additional chemotaxonomic and phenotypic characterisation and deposition in culture collections (Chun et al. 2018).

5.3. Biosynthetic potential associated essentially with the genomic lineage

The near-complete 8.47 Mb draft genome of strain LKIT with 71.6% G+C content and 7,873 predicted protein-coding sequences provides a substantial genomic basis to support the taxonomic and biosynthetic inferences that follow. The identification of 25 BGCs across 15 distinct classes in LKIT indicates it is a biosynthetically rich streptomycete which displays a diverse biosynthetic repertoire encompassing polyketides, nonribosomal peptides, RiPPs, terpenes, siderophores, and other specialised metabolites. Such diversity is consistent with the extensive biosynthetic capacity documented in actinomycete genomes (Nett et al. 2009; Doroghazi and Metcalf 2013), and a large complement of the genome, possessing regulatory and tailoring genes, is also reported to be characteristic of a prolific secondary metabolite producer (Barka et al. 2016; Bentley et al. 2002). However, mere detection of BGCs does not establish expression, metabolite production, or biological activity under the conditions examined. Only a small proportion of the gene clusters is typically expressed under laboratory culture conditions (Nett et al. 2009), which genome mining can prioritise before any chemical work.

When performing comparative BGC analysis of 25 gene cluster families containing an LKIT BGC, 22 were shared with at least one close relative. The remaining three LKIT-exclusive gene-cluster families correspond to short NRPS regions at contig boundaries (Regions 50.1, 51.1 and 53.1), most likely assembly artefacts. Similarly, the nine LKIT strain-specific orthologous clusters within the six-genome comparison encode regulatory, transport

and sulfur-metabolism functions rather than secondary metabolism. These findings place the principal biosynthetic interest at the level of the undescribed genomic lineage, rather than establishing an exceptional repertoire unique to LKIT. An apparent absence of strain-specific biosynthetic gene clusters should be interpreted cautiously, as the closest relatives of LKIT are largely uncharacterised and genomes are poorly studied, and comparisons with a broader range of more distantly related genomes could still uncover cluster content unique to LKIT. Gene-cluster-family analyses are valuable for resolving this distinction because they compare biosynthetic relatedness across genomes rather than relying only on individual database matches (Navarro-Muñoz et al. 2020).

The key biological novelty therefore lies in the previously undescribed genomospecies to which LKIT belongs. Twelve LKIT-containing gene-cluster families have no characterised MIBiG equivalent, and sixteen are absent from the authentically isolated type strain *S. antibioticus* DSM 40234^{T}. This provides biosynthetic differentiation between LKIT and DSM 40234^{T,} which is concordant with the findings of the ANI and dDDH values mentioned previously. However, BGC distribution is additional evidence, not an independent criterion for species delineation. In the same way, the absence of a MIBiG equivalent indicates only that no experimentally characterised reference cluster matches are present in the current database with MIBiG version 4.0 and also does not establish that the encoded products are chemically new. MIBiG is known as a curated repository of experimentally validated BGCs for secondary metabolites, and its coverage necessarily represents only part of bacterial biosynthetic diversity (Zdouc et al. 2025).

Two architecturally unusual clusters merit particular attention: the 51.4 kb hglE-KS–T1PKS–lanthipeptide-class-I hybrid (Region 17.1), which is conserved across all five genomes of the genomospecies, and the 23.4 kb azole-containing RiPP cluster (Region 12.1), which is shared only with NPDCO48228. Neither cluster is unique to LKIT, which supports their interpretation as lineage-associated rather than strain-specific features. Region 17.1 returns a 13% antiSMASH similarity to hexacosalactone A and yet lacks a corresponding MIBiG member within its BiG-SCAPE family. However, the presence of multiple biosynthetic signatures within a single antiSMASH predicted region does not necessarily indicate that they operate as one integrated hybrid pathway, as neighbouring biosynthetic pathways may also be grouped within the same predicted region. Region 17.1, uniting polyketide, heterocyst-glycolipid-type and RiPP machinery in a rarely described combination, constitutes a rational priority for targeted natural-product investigation because of its uncharacterised status and conservation within the LKIT lineage.

5.4. Antibacterial activity and candidate biosynthetic pathways

Antibacterial activity of LKIT provides experimental evidence that complements the genome-mining results. Greatest inhibition was against Gram-positive cocci, *Micrococcus luteus* and a *Staphylococcus aureus* MRSA strain, whereas activity was moderate against *A. baumannii* and weak or undetectable against the other Gram-negative organisms. The absence of inhibition against *Bacillus cereus* further indicates that the activity was not uniformly directed against Gram-positive bacteria. The weak Gram-negative activity is typical of streptomycete metabolites and is usually ascribed to the permeability barrier of the Gram-negative outer membrane.

Several predicted pathways provide promising candidates for further investigation. The first is the antimycin-associated Region 6.2 (NRPS–T1PKS, 52.6 kb), which matches the antimycin A1B cluster completely (BGC0001455) and carries a complete antimycin assembly logic (Section 4.7) (Sandy et al. 2012; Yan et al. 2012).

Antimycins are respiratory chain inhibitors and act on mitochondrial complex III (Huang et al. 2005) and are best known as antifungal and piscicidal agents (Seipke and Hutchings 2013), so their contribution to antibacterial activity may be partial or indirect. The fredericamycin-type T2PKS cluster (Region 3.1, 81% similarity) is a second candidate, since fredericamycin A is active against Gram-positive bacteria (Wendt-Pienkowski et al. 2005; Chen et al. 2008). The ε-poly-L-lysine cluster (Region 11.1, 100%) is a third; this cationic homopolymer has broad-spectrum antimicrobial activity (Yoshida and Nagasawa 2003; Yamanaka et al. 2008), although the weak Gram-negative activity we observed argues against it being the principal effector. Uncharacterised clusters, Region 17.1 among them, may also contribute. The extract may contain multiple active constituents, including products of currently uncharacterised pathways. These assignments rest on sequence alone because activity was scored qualitatively by disc diffusion method, linking any one cluster to it will require bioassay-guided fractionation with LC-MS and, ideally, gene inactivation or heterologous expression.

5.5. Genomic plasticity and resistance-associated functions

The predicted mobile genetic elements provide additional evidence of genome complexity and suggest a moderate degree of genomic plasticity. PHASTER detected three prophage associated regions, one intact (53.2 kb, most similar to *Streptomyces* phage TG1) spanning about 1.3% of the genome (Arndt et al. 2016), and 27 candidate genomic islands via IslandViewer4 (456 ORFs) were identified, with the latter enriched in transposases, phage-related genes, transporters, iron-uptake systems and efflux pumps (Bertelli et al. 2017), consistent with horizontal gene transfer as a driver of genomic and chemical diversity in actinomycetes (Doroghazi and Metcalf 2013; Ziemert et al. 2014). These predictions should be interpreted carefully because composition-based methods can overestimate genomic islands in high-GC genomes. In addition, the largest predicted island (~324 kb) may not represent a single acquired region, as it could result from assembly errors or an extrachromosomal element. CRISPRCasFinder returned 12 candidate arrays across 10 contigs but no *cas* locus (Couvin et al. 2018); orphan arrays of this kind are more likely low-confidence calls and do not establish a functional CRISPR–Cas defence system.

The resistome is consistent with producer self-protection rather than acquired clinical resistance. CARD-RGI identified six determinants: *vanW*, *vanH* and *vanT* (glycopeptide), *rox* and *HelR* (rifamycin) and *otr*(A) (tetracycline). The *vanH*, *vanW* and *vanT* genes form a partial van-type module; in *Streptomyces*, *vanHAX*-type systems remodel the peptidoglycan precursor terminus and confer intrinsic glycopeptide resistance (Hong et al. 2004). More striking is the presence of *rox*, *HelR* and *otr*(A), because none of the 25 clusters is predicted to make a rifamycin- or tetracycline-type compound; these genes may be a legacy of co-occurring producers in the sediment, or self-protection for cryptic pathways, and their function remains to be tested. Roughly 280 transporter-associated proteins, including ABC and efflux systems, add further export capacity that could support self-resistance. Although antibiotic-producing microorganisms often possess resistance mechanisms for self-protection and are well documented, the present analyses cannot determine whether the identified homologues contribute to self-resistance, competition with other microorganisms, or other physiological function (Cundliffe and Demain 2010; D'Costa et al. 2011).

5.6. Ecological significance of recovery from a comparatively underexplored freshwater habitat and limitations of the study

The recovery of LKIT from Loktak Lake sediment expands the genome-based characterisation of cultured actinobacteria from freshwater wetlands, which have received less attention in natural-product bioprospecting than their terrestrial and marine counterparts. A regional investigation of sediments from Tamdil Lake and the Tlawng and Tuirial rivers demonstrated the occurrence of diverse culturable actinobacteria, including antimicrobial *Streptomyces*, supporting freshwater sediments as productive targets for further exploration (Zothanpuia et al. 2018). Thus, the present study contributes to an established but comparatively less developed area of actinobacterial research.

Loktak Lake should not be considered an unexplored microbial environment, as earlier shotgun metagenomic studies reported diverse microbial communities, including Actinobacteria, in phumdi-associated soil and identified functional differences from other aquatic habitats (Puranik et al. 2016). The more specific contribution of the present study is the recovery and genome-based characterisation of a viable *Streptomyces* representative from lake sediment, linking taxonomic identity with antibacterial activity and predicted biosynthetic potential. The cultured strain provides an opportunity for further functional studies beyond what can be achieved through environmental DNA analysis alone.

Functional annotation further identified numerous regulatory, signalling, and transport proteins, nutrient acquisition genes and regions for physiological responsiveness. About 3.8% of genes (271) fell in COG category Q, and the genome encodes some 508 transcriptional regulators, 488 signal-transduction proteins and 14 cytochrome P450s, a regulatory and tailoring capacity in keeping with a large *Streptomyces* genome (Bentley et al. 2002). Several clusters suit a fluctuating sediment habitat: desferrioxamine (*desA-D*) for iron acquisition under limitation (Barona-Gómez et al. 2004), ectoine (Region 16.1) for osmotic stress (Pastor et al. 2010), and a γ-butyrolactone/A-factor-type autoregulator (Region 8.1) that in *Streptomyces* couples morphological differentiation to secondary metabolism (Ohnishi et al. 2005). Loktak's seasonal water-level swings and decaying phumdi biomass make such adaptations plausible, though this remains a hypothesis; because LKIT's clusters are shared with database genomes from other sources, the data do not establish that the lineage is restricted to freshwater.

The taxonomic and functional conclusions of this study are bounded by reference-genome availability, assembly fragmentation, and the scope of experimental validation. The proposed genomospecies is supported by consistent ANI, dDDH, and phylogenomic relationships. However, its distinction from named species depends on the type strain genomes included in the analysis and resolution of inconsistencies among reference records. The study is based on a single isolate from a single sampling event, and the draft assembly (67 contigs) truncates three BGCs at contig boundaries; therefore, long-read or hybrid sequencing will be required to close the genome and resolve the complete boundaries of these clusters. All the biosynthetic predictions are computational and lack complementary expression or chemical validation, and antibacterial activity was qualitatively assessed, rather than quantitatively by MIC determination.

Future work should therefore include genome closure using long-read or hybrid sequencing. Further, it should prioritise LC-MS/MS-based metabolomic profiling combined with bioassay-guided fractionation to identify

antibacterial activity-specific metabolites, the functional validation of the BGC regions 6.2, 3.1, 11.1, and 17.1 through targeted inactivation or heterologous expression, the quantitative MIC determination against the selected ESKAPE pathogens, and the polyphasic characterisation to provide a formal species description.

6. Conclusion

This study identifies *Streptomyces* sp. LKIT, isolated from Loktak Lake sediment, as a cultured representative of a distinct actinobacterium with Gram-positive-active antibacterial properties. Key findings are: (i) ANI, dDDH and phylogenomic evidence place LKIT in an undescribed genomospecies despite 100% 16S rRNA identity to *S. antibioticus* NBRC 12838^T; LKIT was clearly separated from the DSM 40234^T reference genome and all other examined type-strain genomes. It belongs to a coherent but taxonomically unresolved lineage, highlighting the importance of genome-based identification and critical assessment of reference-genome annotations. (ii) The 8.47 Mb draft genome possesses 25 BGCs across 15 classes; (iii) a six-genome BiG-SCAPE comparison shows that the biosynthetic repertoire is conserved across the genomospecies, with twelve LKIT-associated gene-cluster families lacking characterised reference matches, while 16 were absent from the *S. antibioticus* DSM 40234^T genome. This comparative analysis places much of the biosynthetic potential at the lineage rather than the individual-strain level; and (iv) an antimycin-type cluster (Region 6.2), together with fredericamycin-type (Region 3.1) and ε-poly-L-lysine (Region 11.1) clusters, are candidate biosynthetic bases for the observed inhibition of Gram-positive cocci, providing an experimental basis for further investigation, although the responsible metabolites remain unresolved. Together, these findings demonstrate the value of integrating cultivation, genome-based taxonomy, and comparative biosynthetic analysis to investigate actinobacteria from comparatively underexplored freshwater sediments.

Statements and Declarations

Funding

This work was supported by the Department of Biotechnology-Institute of Bioresources and Sustainable Development (DBT-IBSD), India, under the project titled “Bioprospecting microbial diversity in benthic and littoral regions of Loktak Lake for effective management and sustainable development.”

Competing interests

The authors declare no competing interests.

Declaration of Generative AI Use

The language of this manuscript was refined with the assistance of a large language model (LLM), specifically, to improve clarity, grammar, and readability. No generative AI tools were used to create or modify any figures in the manuscript**.**

Author Contributions

Sudhansu Kumar Gouda performed the experiments and analysed the data, including the genome analysis, and wrote the main manuscript text. Khushbu Kumari assisted with sampling, isolation and primary antibacterial screening. Ananta Narayan Panda assisted with the genome analysis of *Streptomyces* sp. LKIT. Vishakha Raina

conceptualized and supervised the study, revised the manuscript and acquired funding along with Dinabandhu Sahoo. All authors read and approved the final manuscript.

Data availability

Streptomyces sp. LKIT has been deposited in the Microbial Type Culture Collection and Gene Bank (MTCC), Chandigarh, India, under accession number MTCC 14074.

The whole genome shotgun project of *Streptomyces* sp. LKIT has been deposited at DDBJ/ENA/GenBank under accession JCERQC000000000 (BioProject PRJNA1535169; BioSample SAMN63566428). The version described in this paper is JCERQC000000000.1. The 16S rRNA gene sequence has been deposited in GenBank under accession number QB063908.

Supplementary Information

Online Resource 1 (SM_1.pdf) Tables S1-S10 and Figs. S1 and S2: genome assembly statistics; full antiSMASH biosynthetic gene cluster inventory; BiG-SCAPE gene-cluster families; genomic islands; CARD-RGI resistance determinants; antimycin cluster gene inventory; genome-wide NRPS/PKS substrate specificity; clinker homology of priority clusters; LKIT-specific orthologous clusters within the six-genome comparison; pairwise ANI matrix and heatmap for the six genomes; BiG-SCAPE similarity network.

References

Alanjary M, Steinke K, Ziemert N (2019) AutoMLST: an automated web server for generating multi-locus species trees highlighting natural product potential. Nucleic Acids Res 47:W276–W282. https://doi.org/10.1093/nar/gkz282

Alcock BP, Huynh W, Chalil R et al (2023) CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res 51:D690–D699. https://doi.org/10.1093/nar/gkac920

Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS (2016) PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44:W16–W21. https://doi.org/10.1093/nar/gkw387

Auch AF, von Jan M, Klenk H-P, Göker M (2010) Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand Genomic Sci 2:117–134. https://doi.org/10.4056/sigs.531120

Aziz RK, Bartels D, Best AA et al (2008) The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75

Bankevich A, Nurk S, Antipov D et al (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021

Barka EA, Vatsa P, Sanchez L et al (2016) Taxonomy, physiology, and natural products of Actinobacteria. Microbiol Mol Biol Rev 80:1–43. https://doi.org/10.1128/mmbr.00019-15

Barona-Gómez F, Wong U, Giannakopulos AE, Derrick PJ, Challis GL (2004) Identification of a cluster of genes that directs desferrioxamine biosynthesis in Streptomyces coelicolor M145. J Am Chem Soc 126:16282–16283. https://doi.org/10.1021/ja045774k

Bauer AW, Kirby WMM, Sherris JC, Turck M (1966) Antibiotic susceptibility testing by a standardized single disk method. Am J Clin Pathol 45:493–496. https://doi.org/10.1093/ajcp/45.4_ts.493

Belknap KC, Park CJ, Barth BM, Andam CP (2020) Genome mining of biosynthetic and chemotherapeutic gene clusters in Streptomyces bacteria. Sci Rep 10:2003. https://doi.org/10.1038/s41598-020-58904-9

Bentley DR, Balasubramanian S, Swerdlow HP et al (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456:53–59. https://doi.org/10.1038/nature07517

Bentley SD, Chater KF, Cerdeño-Tárraga A-M et al (2002) Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). Nature 417:141–147. https://doi.org/10.1038/417141a

Bertelli C, Laird MR, Williams KP, Simon Fraser University Research Computing Group, Lau BY, Hoad G, Winsor GL, Brinkman FSL (2017) IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res 45:W30–W35. https://doi.org/10.1093/nar/gkx343

Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T (2021) antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res 49:W29–W35. https://doi.org/10.1093/nar/gkab335

Blin K, Shaw S, Vader L et al (2025) antiSMASH 8.0: extended gene cluster detection capabilities and analyses of chemistry, enzymology, and regulation. Nucleic Acids Res 53:W32–W38. https://doi.org/10.1093/nar/gkaf334

Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J (2021) eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol 38:5825–5829. https://doi.org/10.1093/molbev/msab293

Cappuccino JG, Sherman N (2005) Microbiology: a laboratory manual, 7th edn. Pearson/Benjamin Cummings, San Francisco

Chater KF (2016) Recent advances in understanding Streptomyces. F1000Res 5:2795. https://doi.org/10.12688/f1000research.9534.1

Chen Y, Wendt-Pienkowski E, Shen B (2008) Identification and utility of FdmR1 as a Streptomyces antibiotic regulatory protein activator for fredericamycin production in Streptomyces griseus ATCC 49344 and heterologous hosts. J Bacteriol 190:5587–5596. https://doi.org/10.1128/jb.00592-08

Chklovski A, Parks DH, Woodcroft BJ, Tyson GW (2023) CheckM2: a rapid, scalable, and accurate tool for assessing microbial genome quality using machine learning. Nat Methods 20:1203–1212. https://doi.org/10.1038/s41592-023-01940-w

Chun J, Oren A, Ventosa A et al (2018) Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. Int J Syst Evol Microbiol 68:461–466. https://doi.org/10.1099/ijsem.0.002516

Couvin D, Bernheim A, Toffano-Nioche C et al (2018) CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res 46:W246–W251. https://doi.org/10.1093/nar/gky425

Cundliffe E, Demain AL (2010) Avoidance of suicide in antibiotic-producing microbes. J Ind Microbiol Biotechnol 37:643–672. https://doi.org/10.1007/s10295-010-0721-x

D'Costa VM, King CE, Kalan L et al (2011) Antibiotic resistance is ancient. Nature 477:457–461. https://doi.org/10.1038/nature10388

De Oliveira DMP, Forde BM, Kidd TJ, Harris PNA, Schembri MA, Beatson SA, Paterson DL, Walker MJ (2020) Antimicrobial resistance in ESKAPE pathogens. Clin Microbiol Rev 33:e00181-19. https://doi.org/10.1128/cmr.00181-19

Doroghazi JR, Metcalf WW (2013) Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes. BMC Genomics 14:611. https://doi.org/10.1186/1471-2164-14-611

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x

Gilchrist CLM, Chooi Y-H (2021) clinker & clustermap.js: automatic generation of gene cluster comparison figures. Bioinformatics 37:2473–2475. https://doi.org/10.1093/bioinformatics/btab007

Guo Z, Ma S, Khan S, Zhu H, Zhang B, Zhang S, Jiao R (2021) Zhaoshumycins A and B, two unprecedented antimycin-type depsipeptides produced by the marine-derived Streptomyces sp. ITBB-ZKa6. Mar Drugs 19:624. https://doi.org/10.3390/md19110624

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Hong HJ, Hutchings MI, Neu JM, Wright GD, Paget MS, Buttner MJ (2004) Characterization of an inducible vancomycin resistance system in Streptomyces coelicolor reveals a novel gene (vanK) required for drug resistance. Mol Microbiol 52:1107–1121. https://doi.org/10.1111/j.1365-2958.2004.04032.x

Huang L-S, Cobessi D, Tung EY, Berry EA (2005) Binding of the respiratory chain inhibitor antimycin to the mitochondrial bc1 complex: a new crystal structure reveals an altered intramolecular hydrogen-bonding pattern. J Mol Biol 351:573–597. https://doi.org/10.1016/j.jmb.2005.05.053

Hutchings MI, Truman AW, Wilkinson B (2019) Antibiotics: past, present and future. Curr Opin Microbiol 51:72–80. https://doi.org/10.1016/j.mib.2019.10.008

Jackman SD, Vandervalk BP, Mohamadi H et al (2017) ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. Genome Res 27:768–777. https://doi.org/10.1101/gr.214346.116

Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S (2018) High-throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114. https://doi.org/10.1038/s41467-018-07641-9

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44:D457–D462. https://doi.org/10.1093/nar/gkv1070

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120. https://doi.org/10.1007/bf01731581

Konstantinidis KT, Tiedje JM (2005) Towards a genome-based taxonomy for prokaryotes. J Bacteriol 187:6258–6264. https://doi.org/10.1128/JB.187.18.6258-6264.2005

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy096

699 Küster E, Williams ST (1964) Selection of media for isolation of streptomycetes. Nature 202:928–929.
700 https://doi.org/10.1038/202928a0

701 Lane DJ (1991) 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M (eds) Nucleic acid techniques in
702 bacterial systematics. Wiley, Chichester, pp 115–175

703 Langille MGI, Hsiao WWL, Brinkman FSL (2010) Detecting genomic islands using bioinformatics approaches.
704 Nat Rev Microbiol 8:373–382. https://doi.org/10.1038/nrmicro2350

705 Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359.
706 https://doi.org/10.1038/nmeth.1923

707 Lewis K (2020) The science of antibiotic discovery. Cell 181:29–45. https://doi.org/10.1016/j.cell.2020.02.056

708 Marmur J (1961) A procedure for the isolation of deoxyribonucleic acid from micro-organisms. J Mol Biol 3:208–
709 218. https://doi.org/10.1016/s0022-2836(61)80047-8

710 Medema MH, Fischbach MA (2015) Computational approaches to natural product discovery. Nat Chem Biol
711 11:639–648. https://doi.org/10.1038/nchembio.1884

712 Medema MH, Kottmann R, Yilmaz P et al (2015) Minimum information about a biosynthetic gene cluster. Nat
713 Chem Biol 11:625–631. https://doi.org/10.1038/nchembio.1890

714 Meier-Kolthoff JP, Göker M (2019) TYGS is an automated high-throughput platform for state-of-the-art genome-
715 based taxonomy. Nat Commun 10:2182. https://doi.org/10.1038/s41467-019-10210-3

716 Naghavi M, Vollset SE, Ikuta KS et al (2024) Global burden of bacterial antimicrobial resistance 1990–2021: a
717 systematic analysis with forecasts to 2050. Lancet 404:1199–1226. https://doi.org/10.1016/s0140-
718 6736(24)01867-1

719 Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW et al (2020) A computational framework to explore large-
720 scale biosynthetic diversity. Nat Chem Biol 16:60–68. https://doi.org/10.1038/s41589-019-0400-9

721 Nett M, Ikeda H, Moore BS (2009) Genomic basis for natural product biosynthetic diversity in the actinomycetes.
722 Nat Prod Rep 26:1362–1384. https://doi.org/10.1039/b817069j

723 Ohnishi Y, Yamazaki H, Kato J-Y, Tomono A, Horinouchi S (2005) AdpA, a central transcriptional regulator in
724 the A-factor regulatory cascade that leads to morphological development and secondary metabolism in
725 Streptomyces griseus. Biosci Biotechnol Biochem 69:431–439. https://doi.org/10.1271/bbb.69.431

726 Park WH (2025) Antimycin A: from mitochondrial poison to multifaceted biological probe and therapeutic agent.
727 Chem Biol Interact 422:111805. https://doi.org/10.1016/j.cbi.2025.111805

728 Parte AC, Sardà Carbasse J, Meier-Kolthoff JP, Reimer LC, Göker M (2020) List of Prokaryotic names with
729 Standing in Nomenclature (LPSN) moves to the DSMZ. Int J Syst Evol Microbiol 70:5607–5612.
730 https://doi.org/10.1099/ijsem.0.004332

731 Pastor JM, Salvador M, Argandoña M et al (2010) Ectoines in cell stress protection: uses and biotechnological
732 production. Biotechnol Adv 28:782–801. https://doi.org/10.1016/j.biotechadv.2010.06.005

733 Puranik S, Pal R, More R, Purohit HJ (2016) Metagenomic approach to characterize soil microbial diversity of
734 Phumdi at Loktak Lake. Water Sci Technol 74:2075–2086. https://doi.org/10.2166/wst.2016.370

735 Richter M, Rosselló-Móra R (2009) Shifting the genomic gold standard for the prokaryotic species definition.
736 Proc Natl Acad Sci USA 106:19126–19131. https://doi.org/10.1073/pnas.0906412106

737 Rong X, Huang Y (2012) Taxonomic evaluation of the Streptomyces hygroscopicus clade using multilocus
738 sequence analysis and DNA–DNA hybridization, validating the MLSA scheme for systematics of the
739 whole genus. Syst Appl Microbiol 35:7–18. https://doi.org/10.1016/j.syapm.2011.10.004

740 Rutledge PJ, Challis GL (2015) Discovery of microbial natural products by activation of silent biosynthetic gene
741 clusters. Nat Rev Microbiol 13:509–523. https://doi.org/10.1038/nrmicro3496

742 Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol
743 Biol Evol 4:406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

744 Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual, 3rd edn. Cold Spring Harbor Laboratory
745 Press, Cold Spring Harbor, NY

Sanasam S, Nimaichand S, Ningthoujam D (2011) Novel bioactive actinomycetes from a niche biotope, Loktak Lake, in Manipur, India. J Pharm Res 4:1707–1710

Sandy M, Rui Z, Gallagher J, Zhang W (2012) Enzymatic synthesis of dilactone scaffold of antimycins. ACS Chem Biol 7:1956–1961. https://doi.org/10.1021/cb300416w

Schubert M, Lindgreen S, Orlando L (2016) AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res Notes 9:88. https://doi.org/10.1186/s13104-016-1900-2

Seemann T (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Seipke RF, Hutchings MI (2013) The regulation and biosynthesis of antimycins. Beilstein J Org Chem 9:2556–2563. https://doi.org/10.3762/bjoc.9.290

Shirling EB, Gottlieb D (1966) Methods for characterization of Streptomyces species. Int J Syst Bacteriol 16:313–340. https://doi.org/10.1099/00207713-16-3-313

Sun J, Lu F, Luo Y, Bie L, Xu L, Wang Y (2023) OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes. Nucleic Acids Res 51:W397–W403. https://doi.org/10.1093/nar/gkad313

Tatusova T, DiCuccio M, Badretdin A et al (2016) NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569

Weisburg WG, Barns SM, Pelletier DA, Lane DJ (1991) 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol 173:697–703. https://doi.org/10.1128/jb.173.2.697-703.1991

Wendt-Pienkowski E, Huang Y, Zhang J et al (2005) Cloning, sequencing, analysis, and heterologous expression of the fredericamycin biosynthetic gene cluster from Streptomyces griseus. J Am Chem Soc 127:16442–16452. https://doi.org/10.1021/ja054376u

Wick RR, Judd LM, Gorrie CL, Holt KE (2017) Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595

Williams ST, Davies FL (1965) Use of antibiotics for selective isolation and enumeration of actinomycetes in soil. J Gen Microbiol 38:251–261. https://doi.org/10.1099/00221287-38-2-251

Williams ST, Goodfellow M, Alderson G, Wellington EMH, Sneath PHA, Sackin MJ (1983) Numerical classification of Streptomyces and related genera. J Gen Microbiol 129:1743–1813. https://doi.org/10.1099/00221287-129-6-1743

Wright MH, Adelskov J, Greene AC (2017) Bacterial DNA extraction using individual enzymes and phenol/chloroform separation. J Microbiol Biol Educ 18:18.2.48. https://doi.org/10.1128/jmbe.v18i2.1348

Xu H, Luo X, Qian J et al (2012) FastUniq: a fast de novo duplicates removal tool for paired short reads. PLoS One 7:e52249. https://doi.org/10.1371/journal.pone.0052249

Yamanaka K, Maruyama C, Takagi H, Hamano Y (2008) ε-Poly-L-lysine dispersity is controlled by a highly unusual nonribosomal peptide synthetase. Nat Chem Biol 4:766–772. https://doi.org/10.1038/nchembio.125

Yan Y, Zhang L, Ito T, Qu X, Asakawa Y, Awakawa T, Abe I, Liu W (2012) Biosynthetic pathway for high structural diversity of a common dilactone core in antimycin production. Org Lett 14:4142–4145. https://doi.org/10.1021/ol301785x

Yoon S-H, Ha S-M, Kwon S, Lim J, Kim Y, Seo H, Chun J (2017a) Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol 67:1613–1617. https://doi.org/10.1099/ijsem.0.001755

Yoon S-H, Ha S-M, Lim J, Kwon S, Chun J (2017b) A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek 110:1281–1286. https://doi.org/10.1007/s10482-017-0844-4

Yoshida T, Nagasawa T (2003) ε-Poly-L-lysine: microbial production, biodegradation and application potential. Appl Microbiol Biotechnol 62:21–26. https://doi.org/10.1007/s00253-003-1312-9

794 Zdouc MM, Blin K, Louwen NL et al (2025) MIBiG 4.0: advancing biosynthetic gene cluster curation through
795 global collaboration. Nucleic Acids Res 53:D678–D690. https://doi.org/10.1093/nar/gkae1115

796 Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome
797 Res 18:821–829. https://doi.org/10.1101/gr.074492.107

798 Zhang W, Che Q, Tan H, Qi X, Li D, Zhu T, Liu M (2023) A novel antimycin analogue antimycin A2c, derived
799 from marine Streptomyces sp., suppresses HeLa cells via disrupting mitochondrial function and depleting
800 HPV oncoproteins E6/E7. Life Sci 330:121998. https://doi.org/10.1016/j.lfs.2023.121998

801 Ziemert N, Lechner A, Wietz M, Millán-Aguiñaga N, Chavarria KL, Jensen PR (2014) Diversity and evolution
802 of secondary metabolism in the marine actinomycete genus Salinispora. Proc Natl Acad Sci USA
803 111:E1130–E1139. https://doi.org/10.1073/pnas.1324161111

804 Zothanpuia, Passari AK, Leo VV, Chandra P, Kumar B, Nayak C, Hashem A, Abd_Allah EF, Alqarawi AA,
805 Singh BP (2018) Bioprospection of actinobacteria derived from freshwater sediments for their potential
806 to produce antimicrobial compounds. Microb Cell Fact 17:68. https://doi.org/10.1186/s12934-018-0912-
807 0

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SM1.pdf