Supplementary Information (Online Resource 1)

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

Journal: Archives of Microbiology

Authors: Sudhansu Kumar Gouda, Khushbu Kumari, Ananta Narayan Panda, Dinabandhu Sahoo, Vishakha Raina*

Corresponding author: Vishakha Raina, vraina@kiitbiotech.ac.in; School of Biotechnology, Kalinga Institute of Industrial Technology (KIIT) Deemed to be University, Bhubaneswar, Odisha 751024, India

Table S1 - Genome assembly, quality and annotation statistics

Feature	Value	Source
Total assembly length	8,469,611 bp (8.47 Mb)	Unicycler v0.4.7 / CheckM2
Number of contigs	67	Unicycler v0.4.7
Longest contig	777,928 bp	QUAST v4.6
N50	327,981 bp	QUAST v4.6
G+C content	71.6%	QUAST v4.6
Genome coverage	~1,018×	Illumina HiSeq 4000 reads
Completeness (CheckM2)	99.99%	CheckM2 (neural-network model)
Contamination (CheckM2)	2.78%	CheckM2
CDS (RAST)	7,873	RAST
CDS (Prodigal/CheckM2)	7,556	CheckM2 QC
tRNA / rRNA / tmRNA	98 / 4 / 1	RAST
Biosynthetic gene clusters	25	antiSMASH 8.0.4
Genomic islands (ORFs)	27 (456 ORFs, 505,044 bp)	IslandViewer 4
CRISPR arrays / cas clusters	12 / 0	CRISPRCasFinder
Resistance determinants (Strict)	6	CARD-RGI

Table S2 - Full biosynthetic gene-cluster inventory (antiSMASH 8.0.4, strict)

Region	Cluster type	kb	Best MIBiG	Nearest cluster	Similarity
1.1	melanin	10.5	BGC0000700.5	*istamycin*	5% · Very low
1.2	NI-siderophore	29.8	BGC0002690.2	*FW0622*	62% · Moderate
2.1	T2PKS; T1PKS	98.8	BGC0000271.5	*spore pigment*	83% · High
2.2	NRPS	44.3	BGC0001574.5	*SF2768*	50% · Moderate
2.3	T1PKS	53.6	BGC0000873.3	*A-201A*	6% · Very low
2.4	terpene; melanin	36.1	BGC0000908.3	*melanin*	71% · Moderate
3.1	T2PKS	72.5	BGC0000224.4	*fredericamycin A*	81% · High
5.1	NRPS	58.9	BGC0000423.5	*scabichelin*	80% · High
6.1	NI-siderophore	31.3	BGC0001732.5	*paulomycin*	11% · Very low
6.2	NRPS; T1PKS	52.6	BGC0001455.6	*antimycin A1B*	100% · High
6.3	indole; hydrogen-cyanide	23.5	BGC0002805.2	*Azodyrecin A/Azodyrecin B/Azodyrecin C*	17% · Very low
8.1	terpene; butyrolactone	22.2	BGC0000850.5	*γ-butyrolactone*	100% · High
9.1	lanthipeptide-class-iii; terpene	35.6	BGC0001456.3	*isorenieratene*	100% · High
11.1	NAPAA	33.9	BGC0002174.2	*ε-Poly-L-lysine*	100% · High
12.1	azole-containing-RiPP	23.4	—	—	— · no/low hit
13.1	furan; butyrolactone	18.8	BGC0000914.5	*methylenomycin A*	14% · Very low
14.1	terpene	21.1	BGC0000660.5	*albaflavenone*	100% · High
16.1	ectoine	10.4	BGC0000853.3	*ectoine*	100% · High
17.1	hglE-KS; lanthipeptide-class-i; T1PKS	51.4	BGC0002497.3	*hexacosalactone A*	13% · Very low
17.2	terpene	26.7	BGC0000663.5	*hopene*	92% · High
25.1	NI-siderophore	30	BGC0000236.5	*kinamycin*	22% · Very low

36.1	NRPS	25.6	BGC0000296.5	*actinomycin D*	28% · Very low
50.1	NRPS	4.7	BGC0002638.3	*misaugamycin A/misaugamycin B*	13% · Very low
51.1	NRPS	4.1	—	—	— · no/low hit
53.1	NRPS	3.1	—	—	— · no/low hit

Tiers: High ≥75%, Moderate 30–74%, Very low <30% or no MIBiG hit. Region 6.2 = antimycin lead; Region 17.1 = priority uncharacterised hybrid.

Table S3 - BiG-SCAPE gene-cluster families containing LKIT BGCs (cutoff 0.3)

Family	LKIT class	Genomes present	In DSM 40234^T	MIBiG member
FAM_00149	*NAPAA*	5 / 6 genomes	no	none
FAM_00141	*NI-siderophore*	6 / 6 genomes	yes	none
FAM_00155	*NI-siderophore*	5 / 6 genomes	no	none
FAM_00143	*NRPS*	5 / 6 genomes	no	none
FAM_00148	*NRPS*	5 / 6 genomes	no	none
FAM_00153	*NRPS*	5 / 6 genomes	no	none
FAM_00157	*NRPS ‡*	1 / 6 genomes	no	none
FAM_00171	*NRPS ‡*	1 / 6 genomes	no	none
FAM_00174	*NRPS ‡*	1 / 6 genomes	no	none
FAM_00137	*T1PKS*	5 / 6 genomes	no	none
FAM_00144	*T1PKS.hglE-KS.lanthipeptide-class-i*	5 / 6 genomes	no	none
FAM_00129	*azole-containing-RiPP*	2 / 6 genomes	no	none
FAM_00135	*butyrolactone.furan*	5 / 6 genomes	no	none
FAM_00130	*hydrogen-cyanide.indole*	5 / 6 genomes	no	none
FAM_00134	*lanthipeptide-class-iii.terpene*	5 / 6 genomes	no	none
FAM_00139	*NI-siderophore*	6 / 6 genomes	yes	present
FAM_00154	*NRPS.T1PKS*	5 / 6 genomes	no	present
FAM_00147	*T1PKS.T2PKS*	6 / 6 genomes	yes	present
FAM_00136	*T2PKS*	5 / 6 genomes	no	present
FAM_00140	*butyrolactone.terpene*	6 / 6 genomes	yes	present
FAM_00138	*ectoine*	6 / 6 genomes	yes	present
FAM_00131	*melanin*	6 / 6 genomes	yes	present
FAM_00142	*melanin.terpene*	6 / 6 genomes	yes	present
FAM_00145	*terpene*	6 / 6 genomes	yes	present
FAM_00150	*terpene*	6 / 6 genomes	yes	present

‡ contig-edge truncated NRPS fragment (assembly artefact). 25 families in total: 22 shared with at least one other genome; 15 with no MIBiG member (12 excluding the three ‡ fragments); 16 absent from authentic S. antibioticus DSM 40234^T.

Table S4 - Genomic islands (IslandViewer 4)

Island	Start	End	Length (bp)	ORFs
GI-01	1,336,898	1,343,034	6,137	3
GI-02	1,483,863	1,488,936	5,074	2
GI-03	1,670,215	1,674,571	4,357	4
GI-04	1,676,381	1,681,155	4,775	3
GI-05	1,849,575	1,858,317	8,743	8
GI-06	2,082,554	2,086,799	4,246	3
GI-07	2,115,658	2,119,763	4,106	6
GI-08	2,167,612	2,172,388	4,777	5
GI-09	3,292,159	3,296,323	4,165	3
GI-10	3,395,950	3,400,595	4,646	3
GI-11	4,609,095	4,618,669	9,575	8

GI-12	5,644,153	5,648,380	4,228	5
GI-13	6,624,408	6,635,756	11,349	15
GI-14	6,674,055	6,678,277	4,223	4
GI-15	7,044,173	7,049,847	5,675	5
GI-16	7,985,946	7,991,908	5,963	4
GI-17	8,080,112	8,404,483	324,372	299
GI-18	8,146,990	8,151,173	4,184	3
GI-19	8,163,638	8,169,669	6,032	3
GI-20	8,196,942	8,203,594	6,653	6
GI-21	8,243,260	8,250,509	7,250	5
GI-22	8,252,494	8,258,140	5,647	7
GI-23	8,275,009	8,281,325	6,317	7
GI-24	8,289,293	8,296,135	6,843	4
GI-25	8,298,041	8,305,423	7,383	6
GI-26	8,313,159	8,320,840	7,682	7
GI-27	8,501,078	8,531,719	30,642	28

27 islands; GI-18 to GI-26 lie within GI-17. Totals are 456 ORFs and 505,044 bp when all 27 islands are summed, or 408 ORFs and 447,053 bp without double-counting the nested islands. Coordinates refer to the concatenated contig order used by IslandViewer 4 and can therefore exceed the assembly length (8,469,611 bp).

Table S5 - Antimicrobial-resistance determinants (CARD-RGI, Strict)

Best-hit ARO	% id	% Coverage	Drug class	Mechanism / interpretation
rox (S. venezuelae)	78.7	100.2	Rifamycin	Rifampin monooxygenase
helR	77.5	99.3	Rifamycin	RNA-pol protection
otr(A) (S. rimosus)	51.8	98.3	Tetracycline	Ribosomal protection
vanH (vanF cluster)	37.5	102.8	Glycopeptide	van homolog (intrinsic)
vanT (vanG cluster)	30.8	60.3	Glycopeptide	van homolog (intrinsic)
vanW (vanI cluster)	29.2	153.9	Glycopeptide	van homolog (intrinsic)

rox and helR are high-identity hits (~78%); otr(A) is a moderate-identity hit (51.8%). The van homologs (29–38% identity, partial/over-length coverage) are intrinsic Streptomyces paralogs, not acquired resistance.

Table S6 - Antimycin-type cluster (Region 6.2): selected genes

Location (bp)	Str	Product (RAST)	Biosynthetic role
5,226–6,510	+	Putative aminotransferase	Tailoring/redox
17,371–17,893	-	RNA polymerase sigma factor RpoE	Regulation
20,000–28,730	+	Polyketide synthase modules and related proteins	Core NRPS–PKS
28,722–32,571	+	Polyketide synthase modules and related proteins	Core NRPS–PKS
32,629–33,853	+	Crotonyl-CoA carboxylase/reductase, ethylmalonyl-CoA producing	Acyl side-chain (antE)
33,946–35,431	-	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3)	Acyl side-chain (antF)
35,925–36,933	+	1,2-phenylacetyl-CoA epoxidase, subunit A (EC 1.14.13.149)	3-Formamidosalicylate head
36,929–37,271	+	1,2-phenylacetyl-CoA epoxidase, subunit B (EC 1.14.13.149)	3-Formamidosalicylate head
37,245–38,031	+	1,2-phenylacetyl-CoA epoxidase, subunit C (EC 1.14.13.149)	3-Formamidosalicylate head
38,027–38,525	+	Phenylacetate-CoA oxygenase, PaaJ subunit	3-Formamidosalicylate head
39,679–40,459	+	Oxidoreductase, short-chain dehydrogenase/reductase family	Tailoring/redox
40,525–41,374	+	Tryptophan 2,3-dioxygenase (EC 1.13.11.11)	3-Formamidosalicylate head
42,211–43,471	+	Kynureninase (EC 3.7.1.3)	3-Formamidosalicylate head

100% MIBiG similarity to antimycin A1B (BGC0001455). Threonine-loading NRPS + malonyl-CoA PKS + thioesterase; 3-formamidosalicylate head from the tryptophan/kynurenine route; crotonyl-CoA carboxylase/reductase supplies the variable acyl chain.

Table S7 - Genome-wide NRPS/PKS substrate specificity

Region	Cluster class	Predicted monomers / extender units
6.2	antimycin-type (NRPS–PKS)	A: Thr, X · AT: malonyl-CoA · trans-A: X
5.1	scabichelin-type siderophore	A: N5-formyl-N5-OH-Orn, Ser, OH-Orn, Orn, OH-Orn
2.1	spore-pigment T1PKS–T2PKS	AT: methylmalonyl- and malonyl-CoA
3.1	fredericamycin-type T2PKS	AT: malonyl-CoA · A: X
36.1	actinomycin-type NRPS	A: X, Thr, Val
17.1	orphan hglE-KS–lanthipeptide–T1PKS	AT: malonyl-CoA, pk
50.1 / 51.1 / 53.1	NRPS fragments (contig edges)	A: Pro / Val / Gly

Predictions from antiSMASH 8.0.4 NRPS/PKS domain analysis. A, adenylation domain; AT, acyltransferase domain; trans-A, adenylation domain acting in trans; X, no confident substrate prediction; pk, polyketide extender unit not assigned.

Table S8 - Clinker homology of priority clusters

Region	Comparison	Shared genes (≥50% id)	Status
17.1 (hglE-KS/T1PKS/lanthipeptide-I)	LKIT vs NPDCO48228	31	conserved
17.1	LKIT vs DSM 41481	25	conserved
17.1	LKIT vs Amel2xC10	25	conserved
17.1	LKIT vs NRRL B-1701	25	conserved
12.1 (azole-RiPP)	LKIT vs NPDCO48228	21	conserved
12.1	LKIT vs Amel2xC10	14	partial (not in same GCF)

Region 17.1 is conserved across the genomospecies; Region 12.1 is shared with NPDCO48228 and only partially with Amel2xC10 (not grouped in the same BiG-SCAPE family). Neither cluster is strain-unique, corroborating the BiG-SCAPE result. Quotation marks denote a genome deposited under a type-strain number whose identity is not supported by whole-genome comparison.

Table S9 - LKIT species-specific orthologous clusters (OrthoVenn3, six-genome panel)

Cluster ID	# proteins	Swiss-Prot hit	Function (GO annotation)
cluster6768	3	N/A	*Unannotated*
cluster6769	3	Q845S7	Transcription-factor binding (GO:0008134)
cluster6770	3	P46068	Transcription, DNA-templated (GO:0006351)
cluster7310	2	P33950	Dipeptide transport (GO:0042938)
cluster7311	2	N/A	*Unannotated*
cluster7312	2	C1DN48	Alkanesulfonate monooxygenase (GO:0008726; sulfur metabolism)
cluster7313	2	P76462	Putative permease (no GO term)
cluster7314	2	N/A	*Unannotated*
cluster7315	2	P37409	Ion transport (GO:0006811)

None of the nine LKIT-specific clusters encodes a biosynthetic function, corroborating the BiG-SCAPE result that no biosynthetic gene cluster is unique to LKIT. Functions are limited to regulation, transport and sulfur metabolism, with three clusters unannotated.

Table S10 - Pairwise average nucleotide identity (FastANI, %)

Genome	LKIT	NPDCO48228	"NRRL B-1701"	DSM 41481	Amel2xC10	DSM 40234^{T}
LKIT	100.0	99.0	96.9	96.8	96.4	85.3
NPDCO48228	99.0	100.0	96.9	96.9	96.5	85.2
NRRL B-1701	96.9	96.9	100.0	98.8	97.3	85.2
DSM 41481	96.9	97.0	98.9	100.0	97.2	85.4
Amel2xC10	96.5	96.5	97.3	97.2	100.0	85.4
DSM 40234^{T}	85.3	85.4	85.4	85.4	85.3	100.0

FastANI values (%) (Jain et al. 2018); values are directional (row = query, column = reference). The five genomes LKIT, NPDCO48228, NRRL B-1701, DSM 41481 and Amel2xC10 form a single genomospecies (≥96% ANI to one another). The authentic S. antibioticus type strain DSM 40234^{T} is ~85% ANI from all five (values in bold), i.e. a different species. Genuine subcultures of one type strain would be ≥99% identical; the ~85% identity between DSM 40234^{T} and the genomes deposited as NRRL B-1701, NPDCO48228 and DSM 41481 indicates that these deposits are putatively misidentified and conspecific with LKIT. Quotation marks as in Table S8; ^{T}, type strain.

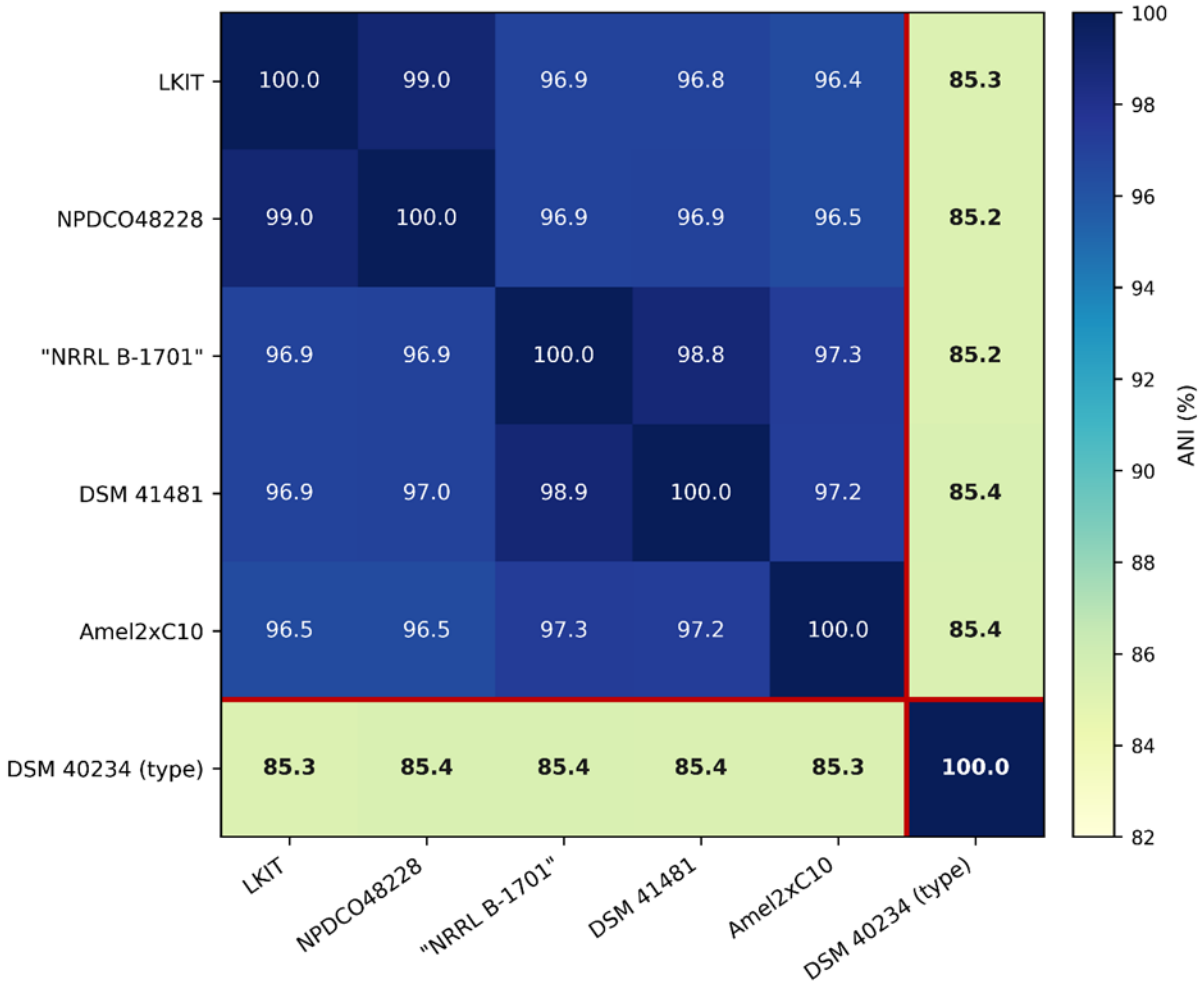


Fig. S1 Heatmap of the Table S10 values. Colour scales with ANI; the red lines separate the authentic *S. antibioticus* type strain DSM 40234^{T} (uniformly ~85% ANI, pale cells) from the LKIT genomospecies (≥96% ANI, dark cells)

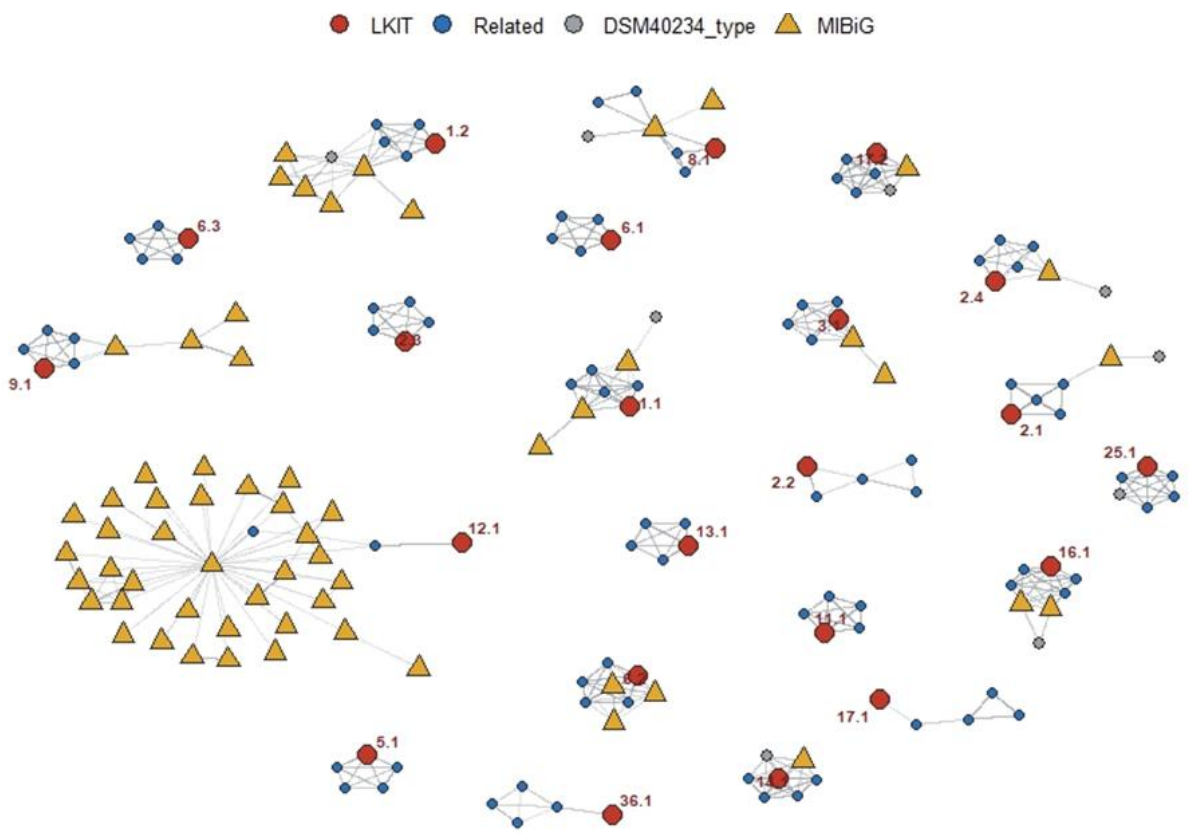


Fig. S2 Similarity network of the gene-cluster families containing *Streptomyces* sp. LKIT biosynthetic gene clusters, from the six-genome BiG-SCAPE comparison (mix mode, distance ≤ 0.30). Nodes are gene clusters: red, LKIT (labelled by antiSMASH region); blue, clusters from the related genomes; grey, authentic *S. antibioticus* type strain DSM 40234^{T}; gold triangles, MIBiG reference clusters. Edges link clusters with BiG-SCAPE distance ≤ 0.30 (edge opacity scales with similarity). A red LKIT node grouped with a gold triangle corresponds to a family with a characterised MIBiG match; families without a MIBiG node are uncharacterised. The three contig-edge NRPS fragments (Regions 50.1, 51.1, 53.1) are singletons and are not shown. This network is the raw similarity view; definitive family assignments are given in Table S3