ORIGINAL ARTICLE

Draft genome sequence of *Streptomyces* sp. KD18, isolated from industrial soil

Khushboo[1] · Nirjara Singhvi[2] · Vipin Gupta[3] · Namrata Dhaka[1] · Kashyap Kumar Dubey[4]

Abstract

The present study scrutinizes the presence of *Streptomyces* strains in the soil sample collected from industrial area of Bahadurgarh (Haryana) India. The morphological approach manifested the isolated strain belong to *Streptomyces* species and named as *Streptomyces* sp. KD18. Sequencing of *Streptomyces* sp. KD18 genome was performed by Illumina Nextseq500 platform. 65 contigs were generated via SPAdes v3.11.1 and harboured genome size of 7.2 Mb. AntiSMASH server revealed the presence of 25 biosynthetic gene clusters in KD18 genome where BGC of lipstatin was of more interest from industrial and pharmaceutical purpose. The draft genome sequence represented via ANI values claimed that the KD18 strain belongs to *Streptomyces toxytricini* and finally named as *S. toxytricini* KD18. The LC–MS analysis of the extracted metabolite confirmed the production of lipstatin. The genome sequence data have been deposited to NCBI under the accession number of *GCA_014748315.1*.

Keywords *Streptomyces* · Lipstatin · Genomics · Pan genome · High-throughput genomics

✉ Kashyap Kumar Dubey
kashyapdubey@gmail.com

Khushboo
khusi079@gmail.com

Nirjara Singhvi
nirjara20@gmail.com

Vipin Gupta
guptavipin142@gmail.com

Namrata Dhaka
namrata@cuh.ac.in

1 Department of Biotechnology, Central University of Haryana, Mahendergarh 123031, India

2 School of Allied Sciences, Dev Bhoomi Uttarakhand University, Dehradun, Uttarakhand 248007, India

3 Ministry of Environment, Forest and Climate Change, Integrated Regional Office, Dehradun, Uttarakhand 248001, India

4 Bioprocess Engineering Laboratory, School of Biotechnology, Jawaharlal Nehru University, New Delhi 110067, India

Introduction

Actinobacteria are copious in nature, with nearly 500 formally illustrated species (Goodfellow 2012; Labeda et al. 2012). The actinobacterial species possess a complex life cycle along with the production of diverse pharmaceutically important compounds (Claessen et al. 2014). More than 10,000 bioactive metabolites, such as anti-biotics, anti-cancer, anti-tumor, anti-obesity, anti-fungal, anti-microbial, anti-viral, volatile organic compounds, and so on, are synthesised by *Streptomyces* species, and many more bioactive compounds must be identified for both economic and health benefits (Zerouki et al. 2021). Genome sequencing and bioinformatics provide a better understanding of how gene sequences synthesize the functional small molecules that are important for life. This revolution has facilitated the discovery of a vast array of natural products, many of which have been found to possess novel therapeutic activities. Biosynthetic Gene Clusters (BGCs) and a huge diversity of polyketides and non-ribosomal peptides have been identified from diverse bacteria and fungi (Majer et al. 2021). BGCs represent the group of genes involved in the metabolism, i.e., the production, transport and regulation of secondary metabolites (Williams 2013; Beld et al. 2014; Luo et al. 2014; Baltz 2019; Galanie et al. 2020).

Springer

The genus *Streptomyces* has been extensively studied in recent years in search of new bioactive compounds. The goal of the current investigation was to determine the taxonomic status of the isolated *Streptomyces* sp. KD18 followed by the investigation of its potential to produce bioactive compounds. For this, the isolated strain KD18 was subjected to genome sequencing and analysis. In addition, liquid chromatography–mass spectrometry (LC–MS) was performed to identify the secondary metabolites in the fermentation liquid, confirming the presence of lipstatin BGC. The genome sequence of *Streptomyces* sp. KD18 might be utilized to better understand the genetic basis involved in the production of secondary metabolites along with future bioengineering purposes.

Materials and methods

Sampling source and strain isolation

Samples were collected from many diverse habitats (forest area, agriculture land, industrial area) from Bahadurgarh, Haryana, India. The colonies were grown in inorganic salt starch agar (ISSA; ISP medium No. 4, HiMedia M359, India), and starch casein agar medium (HiMedia M2054, India) as a procedure for selective isolation of *Streptomyces* (Amin et al. 2017). Isolated culture was sub-cultured in 30 ml of YEME broth medium (yeast extract 4.0 g/L, malt extract 10.0 g/L, dextrose 4.0 g/L, $CaCO_3$ 2.0 g/L) containing 0.5% glycine for 48 h at 28 °C and 220 rpm in shaking incubator. Isolation of genomic DNA was executed as described by Nikodinovic and group, 2003 (Nikodinovic et al. 2003).

Genome sequencing, assembly and annotation

Paired end sequencing library was constructed from the isolated genomic DNA using Illumina library preparation kit. The genome was sequenced with Illimina-Nextseq500 sequencing platform 150 bp X 2 paired end read length. The genomic high-quality reads were assembled and validated using SPAdes version 3.11.1 (Bankevich et al. 2012). SSPACE version 3.0 (Boetzer et al. 2011). Genome was annotated using NCBI PGAP pipeline, PROKKA v1.12 and RAST server based on the RAST tool kit (RASTtk). CRISPR finder was also used to annotate CRISPRs in the genome. For further analysis, AntiSMASH 6 beta bacterial version webserver was used for the identification of BGCs (Medema et al. 2011; Seemann 2014; Aziz et al. 2008). AMPHORA2 was used to identify the presence of single copy essential genes (Wu and Scott 2012)in the KD18 genome.

Phylogeny and pathway mapping

Whole genome-based phylogenetic analysis was carried out for 10 *Streptomyces* strains. The phylogenetic tree was constructed using maximum likelihood methodology with 1000 bootstrap value and visualized using iTOL (Letunic and Bork 2021). Average Nucleotide Identity (ANI) was calculated using the ANI matrix of Kostas lab (Goris et al. 2007)for the taxonomic demarcation of strain KD18 and 9 reference genotypes and visualized using Morpheus Broad Institute's webserver (Morpheus, https://software.broadinstitute.org/morpheus). M1CR0B1AL1Z3R webserver was used for core genome length and ORF counting of all the selected genomes (Avram et al. 2019).

Core-pan genome analysis

Bacterial pan genome analysis (BPGA) pipeline was used to examine the core and accessory genome of the selected strains (http://www.iicb.res.in/bpga/index.html) (Chaudhari et al. 2016) whereas Othro Venn software was used for the alignment and prediction of highly conserved proteins among the 10 selected strains (http://www.bioinfogenome.net/OrthoVenn/start.php) (Wang et al. 2015).

Extraction and analysis of lipstatin

For the production of lipstatin, the bacteria were inoculated in production medium containing soy flour 35.0 g/L, glycerol 22.5 ml/L, soy oil 25.0 ml/L, soy lecithin 15.0 g/L, polypropylene glycol 0.5 g/L and grown at 28 °C with 220 rpm for 10 days. The extraction and HPLC analysis of lipstatin was carried out as described by Kumar and group, 2012 (Kumar et al. 2012).The crude extract of the secondary metabolite produced by *Streptomyces* sp. KD18 was subjected to LCMS analysis and followed the methodology as described by Yang and group, 2019 (Yang et al. 2019).

Results and discussion

Strain culturing and growth patterns

The cultured strains were analyzed morphologically and the representative colonies were further inoculated on new culture plates of ISSA and starch casein agar medium. Storage of selected colonies were done on Streptomyces agar medium (HiMedia M1352, India). The *Streptomyces* sp. KD18 thrived on inorganic salt starch agar (ISSA) and casein agar medium. Initially, the colonies were opaque white with dry surface and rough texture. After 6–7 days, pink-colored

colonies were clearly visible with earthy odor and long-growing hyphae were observed under microscope without any fracture over surface. Pellet formation was observed in the submerged culture. Morphological analysis revealed that the isolated KD18 strain was related to *Streptomyces* sp. The selected isolates were used for molecular analysis.

General genomics attributes

Using Illumina NextSeq 500 sequencing technology, a total of 4,924,225 paired-end reads were generated from a single library. After low-quality reads were removed, 48,748,878 (99.00%) high-quality reads were used for de novo genome assembly. A total of 65 scaffolds were created with N50 value of 190,392 bp.

Genome annotation

Annotation using NCBI (prokaryotic genome annotation pipeline) PGAP

The size of *Streptomyces* sp. KD18 genome was 7.2 Mb with 73.8% G + C content. It has been illustrated earlier that the genomes with high G + C content are generally large and possess a large number of open reading frames (ORFs) (Wu et al. 2012; Boldögkoi et al. 1995). The genome of *Streptomyces* sp. KD18 was annotated using the PGAP workflow. The genome harboured 6462 genes. Nearly 6235 genes were coding for the proteins. With 0.71 GC content, the genome of *Streptomyces* sp. KD18 contained 8791 ORFs.

Out of 6235 annotated CDSs, only 5163 were grouped into the pre-existing COG categories. It was observed that most of these genes encoded hypothetical proteins, i.e., the proteins with unknown function. The proteins with unknown function might be the novel proteins whose function cannot be predicted using existing homology-based annotation techniques.

Annotation via PROKKA software

5 rRNA and 94 tRNA were predicted using PROKKA software. Average size of genes was 958 bp. Maximum size of gene was 14,256 bp while the gene of minimum size was of 59 bp.

RAST server: rapid annotation using subsystems technology

According to RAST analysis, 181,447 bp N50 and 13 L50 were predicted. In addition, 309 subsystems were identified using RAST annotation. *S. avermitilis* MA-4680, *S. scabiei* 87.22, *S. coelicolor* A3(2) and *S. griseus subsp. griseus* NBRC 13,350 with 542, 520, 489 and 430 score,

respectively, were identified as the closest neighbours of *S.* sp. KD18. The four identified neighbours were selected to examine the evolutionary relationship of KD18 strain. The KD18 strain was found to contain the genes belonging to stress response, iron acquisition and metabolism, nitrogen metabolism, phosphorous metabolism, etc. The KD18 strain possessed subsystem of about 374 for amino acids and derivatives, 258 for carbohydrates, 231 for protein metabolism and 52 for stress response (Fig. 1).

Functional annotation via KEGG

To identify the potential involvement of predicted CDSs of *Streptomyces* genome in the biological pathways, the genes were mapped to the canonical pathways in KEGG. All the genes were classified in five different categories such as metabolism, cellular Processes, genetic information processing, environmental information processing and organismal system. Output of the KEGG automated annotation server (KAAS) includes KEGG orthology assignments with corresponding enzyme commission numbers and metabolic pathways of predicted genes. 6543 predicted genes of KD18 genome were processed in KAAS. A total of 1952 genes were involved in 24 KEGG pathways. Metabolism process contains the highest gene count, i.e., 1362. In metabolism category, amino acid metabolism has the highest gene counts, i.e., 279, followed by 257 and 159 in "carbohydrates metabolism" and "metabolism of cofactors and vitamins", respectively, whereas environmental adaptation pathways of organismal system have the lowest gene count, i.e., 11.

Secondary metabolites gene clusters

The biosynthetic gene cluster analysis revealed the presence of 25 BGCs of different secondary metabolites in *Streptomyces* sp. KD 18 (Table 1). The identified BGCs included non-ribosomal peptide synthase (NRPS) (4), T3PKS (1), T1PKS (1), siderophore (3), terpene (6), butyrolactone (1), RiPP-like, lanthipeptide (2), thiopeptide, CDPS. Out of 25 predicted BGCs, 6 clusters exhibited 90% or above similarity with already known BGCs, suggesting that all the genes required for the synthesis of respective compound are present and organism can produce these metabolites under suitable culture conditions. 4 clusters showed above 50% similarity with respect to BGCs data deposited in MIBiG database. Eleven BGCs in the KD18 genome assembly exhibited low sequence similarity, i.e., between 4 and 30%, while four BGCs displayed null sequence identity with previously described BGCs, indicating the potential to produce new natural compounds. All these identified clusters possess various clinically important applications.

The percentage similarity is a measure of how closely related a given BGC is to the reference cluster on MIBiG

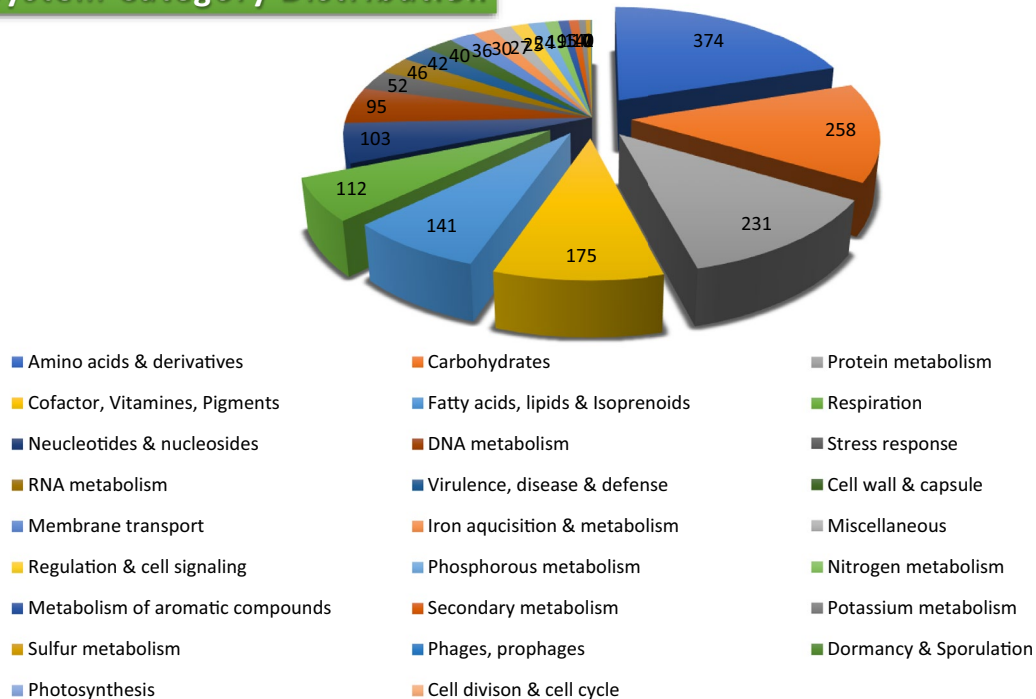## Subsystem Category Distribution

Fig. 1 An overview of subsystems with their associated genes present in KD18 genome annotated using RAST v5.0

database. Notably, the non-ribosomal peptidase synthase (NRPS) cluster with 92% similarity to the known BGC of lipstatin was observed in *S.* sp. KD18 and *S. toxytricini*. Along with *S. toxytricini, S. virginiae* is also the producer of lipstatin. According to the NPATLAS database entry, it is a linear polyketide as well as an organic compound derived from leucine or its derivates. The compound acts as an irreversible inhibitor of pancreatic lipase. Furthermore, the β-lactone ring also inhibits the fatty acid synthase (FAS) activity via binding to the thioesterase domain and exhibits anticancer activity (Mulzer et al. 2014).

Lanthipeptide-class-iv gene cluster (contig 22) was found to demonstrate 100% similarity to the venezuelin BGC in *Streptomyces venezuelae* ATCC 10712 (Goto et al. 2010). Lanthibiotic synthetases are the most important biocatalysts as they use only two posttranslational modification reactions, *i.e.,* Ser/Thr residues dehydration and Cys thiols addition to the resulted amino acids to form conformationally amenable peptides to execute different biological functions (Goto et al. 2010).

Siderophore gene cluster (contig 25) was showing 100% similarity with the desferrioxamine B BGC in *Streptomyces griseus subsp. griseus* NBRC 13350. Desferrioxamine B is marketed under the trade name "Desferal" and possess the clinical history associated with chronic iron overload due to β-thalassemia (Telfer et al. 2019).

A gene cluster present at contig 14 was identified as nucleoside BGCs with 100% similarity to tunicamycin B1 BGC in *Streptomyces chartreusis* NRRL 3882 (Wyszynski et al. 2012). *Streptomyces chartreusis, Streptomyces lysosuperificus* are the main producers of Tunicamycin B1 (Shiraishi et al. 2005). Tunicamycins are the nucleoside antibiotics specifically targeting the MarY enzyme of bacteria to inhibit the synthesis of lipid I precursors and thus inhibit the cell wall formation (Bible and Ryder 2016). In addition, these antibiotics also restrict *N*-linked glycosylation in eukaryotes which leads to arrest of cell cycle at G1 phase. The antibiotic is used to induce the response to unfolded proteins (Chen et al. 2009). Tunicamycin also possesses the ability to cause redifferentiation in anaplastic thyroid cancers (ATCs) via inducing the restoration of genes involved in iodide handling and radioiodine avidity during radioiodine therapy (Choi et al. 2021). Terpene type BGC (contig 23) was showing 100% similarity with isorenieratene biosynthetic gene cluster in *Streptomyces griseus subsp. griseus* NBRC 13350. *Streptomyces* genomes, including *Streptomyces* sp. KD18, are also predicted to contain a core set of natural products encoding genes like desferrioxamine B, hopene along with some common clusters such as ectoine, melanin, isorenieratene, etc. (Kim et al. 2015; Seipke 2015; Komaki et al. 2018; Vicente et al. 2018; Benaud et al. 2021). Through further research, a better understanding of the relationship

Table 1 Biosynthetic gene clusters identified across *Streptomyces* sp. KD18 genome

Contig no	Size (nt.)	Similar biosynthetic cluster	% Similarity	MIBiG BGC-ID
RiPP				
1	20,546	Legonaridin	16	BGC0001188
22	22,674	Venezuelin	100	BGC0000563
24	34,722	Lactazole	66	BGC0000606
12	11,578	–	0	–
NRP				
2	204,321	Lipstatin	92	BGC0000382
12	67,001	Ashimides	12	BGC0001961
13	59,217	Caperomycin	9	BGC0000316
34	44,392	Salinichelins	69	BGC0001767
NRP + RiPP				
2	74,354	Pheganomycin	57	BGC0001148
Siderophore				
3	13,227	–	0	–
11	12,115	–	0	–
Saccharide				
35	29,631	Acarbose	10	BGC0000691
Polyketide				
11	30,658	Granaticin	8	BGC0000227
21	21,482	Ebelactone	5	BGC0001580
30	38,559	Monensin	26	BGC0000100
Terpene				
18	46,558	Hopene	30	BGC0000663
44	14,724	Hopene	30	BGC0000663
22	17,096	–	0	–
24	26,103	Isorenieratene	100	BGC0000664
CDPS				
4	11,257	–	0	–
Other				
3	45,166	Cystargolide	50	BGC0000880
4 (Nucleoside)	28,750	Tunicamycin B1	100	BGC0001440
25	11,785	Desferrioxamin B	100	BGC0000941
43	22,217	Toxoflavin	14	BGC0001972
3	40,687	Melanin	100	BGC0000911

Percentage similarity with already known clusters and their MIBiG IDs are also mentioned

between the genome completeness and AntiSMASH output will be achieved.

Phylogenetic analysis

Even though a single genome analysis can yield useful information, comparative examination of many genomes adds a whole new level of biological knowledge and can produce unexpectedly important findings (Ikeda et al. 2003). Four closest neighbours, i.e., *S. avermitilis* MA-4680, *S. scabiei* 87.22, *S. coelicolor* A3(2) and *S. griseus subsp. griseus* NBRC 13350 were extracted from RAST while four other members i.e., *S. globosus* LZH-48, *S. scabiei* 87.22, *S.* sp. fd1-xmd, *S.* sp. TN58 were selected from BLAST analysis.

Due to 16S rRNA gene sequence similarity, *S. toxytricini* was also selected for comparative and phylogenetic analysis. Among the 10 selected genomes, *S. scabiei* 87.22 had the largest genome with a total length of 10,148,695 bp and the GC content ranged from 69.9 to 73.9%. Details of each genome are described in Table S1.

Total 1306 genes were predicted as core genes and were aligned using muscle aligner to reconstruct phylogeny. It was observed that *S.* sp. KD18 shared the clade with *S. globus* and *S. toxytricini* (Fig. 2A). Similarly, on the basis of pairwise comparison, the ANI have clustered strain KD18 and *S. toxytricini* in a monophyletic clade with *S. globous. S. coelicolor* and *S. scabiei* were clustered together according to ANI topology (Fig. 2B). Thus, with multiple phylogenetic
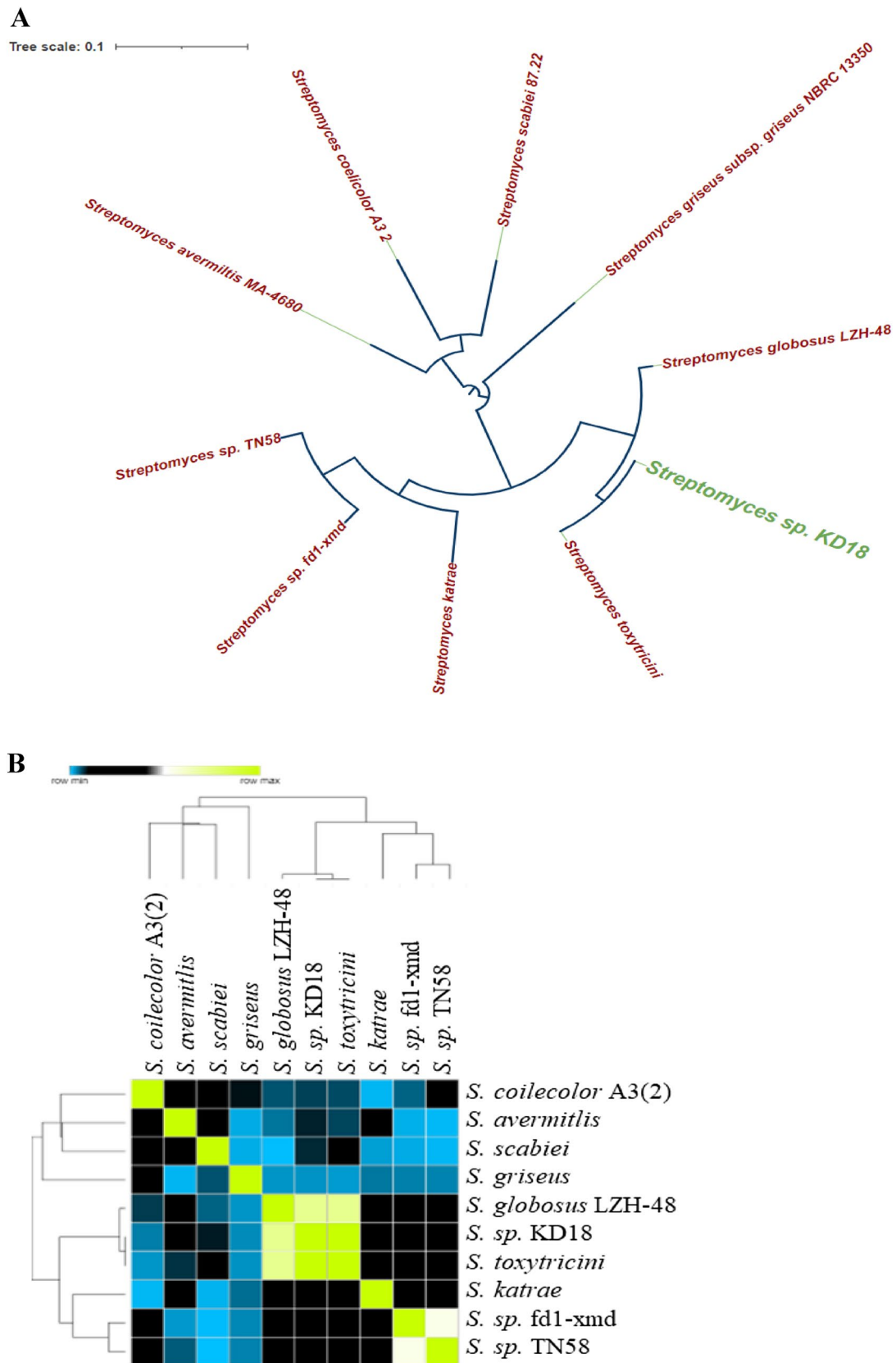
A

Tree scale: 0.1

B

Fig. 2 Phylogenetic analysis of *Streptomyces* sp. KD18. **A** ML-based phylogenomic tree construction based on whole genome sequence of the 10 selected strains of *Streptomyces* used in this study. **B** Correlation plot for *Streptomyces* sp. KD18 and its 9 reference genotypes on the basis on the basis of pairwise ANI values Maximum (green: ANI) and minimum (blue: ANI) correlation are shown with the color scale

analysis it was predicted that the strain KD18 was closest to the species *Streptomyces toxytricini*. *Streptomyces* sp. KD18 and *S. toxytricini* JCM were 100% similar at ANI level. Therefore, on the basis of ANI threshold related to sub-species demarcation (> 98%) (Konstantinidis and Tiedje 2005), it was concluded that the isolated strain KD18 can be considered as sub-species and should be classified as *S. toxytricini* KD18.

Core-pan genome prediction

From the Core-Pan plot, it was clear that the core genes were decreasing gradually along with the addition of new genomes. It was observed that the pan genome was composed of 22,218 genes, out of which only 2341 genes were conserved among the selected species as the core genome. Maximum number of accessory genes (3710) were present in *S. toxytricini* while *S. griseus* possessed minimum number of accessory genes (2349). In case of unique genes, *Streptomyces scabiei* 87.22 (2507) was at the top while *Streptomyces* sp. KD18 was having only 21 unique genes. An open pan genome was observed for the *Streptomyces* sp., which suggests the addition of new genes in the accessory genome of newly sequenced genomes.

OrthoVenn software was used for the further determination of essential genes. A total of 10,703 protein clusters were identified with 8119 orthologous clusters and 2584 single-copy gene clusters, from the 10 selected species of *Streptomyces*. The number of protein clusters for each species varied: *Streptomyces* sp. KD18 (6134), *S. avermitilis* (6059), *S. scabiei* (6364), *S. coelicolor* A3(2) (6103) and *S. griseus* (5641) *S. globosus* (5754), *S.* sp. fd1-xmd (6023), *S.* sp. TN58 (5843), *S. toxytricini* (6221) and *S. katrae* (5646). The result indicated that 3056 orthologous genes were shared by six selected genomes (Fig. 3). The COG distribution analysis revealed that the main functions of the core genes were associated with central dogma, production and conversion of energy along with primary metabolic processes, whereas the secondary metabolic processes were mainly under the control of accessory and unique genes instead of the core genes.

The KEGG analysis disclosed that a large proportion of the core, accessory as well as unique genes controls the carbohydrate, amino acid and lipid metabolism processes. In the categories of membrane transport and metabolism of terpenoids and polyketides, the accessory and unique genes are the key players. COGs distribution revealed that core genes were associated with central dogma, production and conversion of energy along with primary metabolic processes whereas the secondary metabolic processes were mainly under the control of accessory and unique genes instead of the core genes.

Fig. 3 Venn diagram generated through orthoVenn showing unique and shared protein clusters among six selected taxa
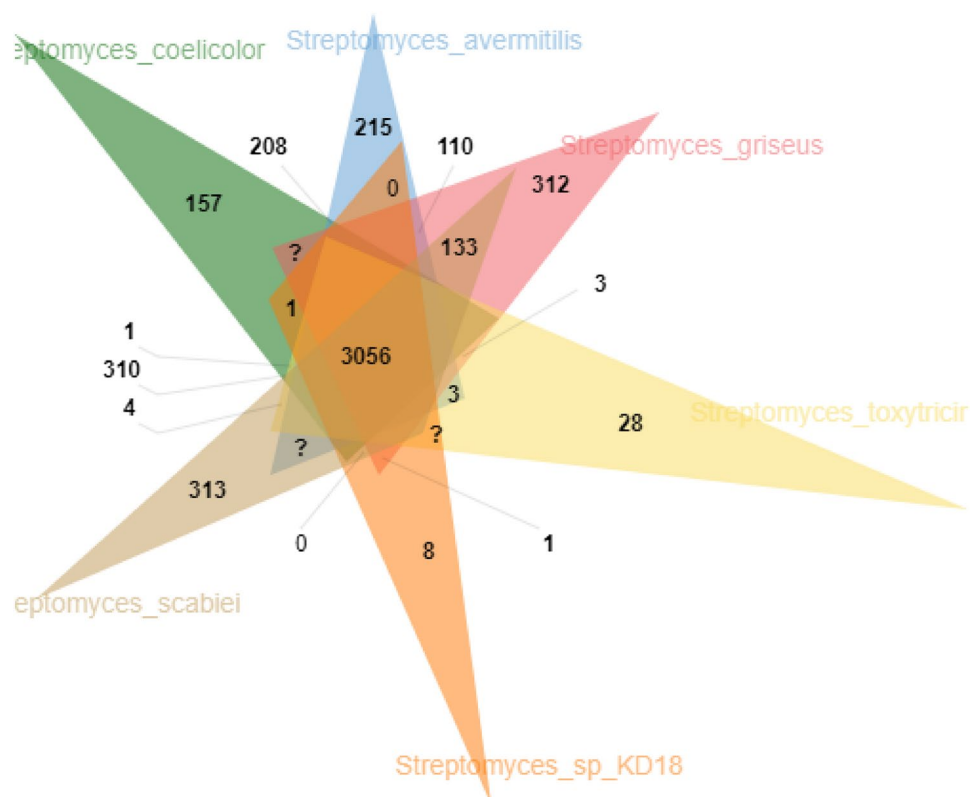
Production of lipstatin

As the BGC of lipstatin was identified through antiSMASH analysis, the *Streptomyces* sp. KD18 culture was subjected to the production of lipstatin. The UV pattern of the crude extract was similar to that of pure lipstatin, where the maximum absorbance was observed at 210 nm. Through HPLC analysis, it was observed that 2.06 mg/ml of lipstatin was produced by the KD18 strain. The retention time for the lipstatin was 1.45 min. All these results confirmed the production of lipstatin by the isolated strain KD18.

LC–MS analysis of crude lipstatin

Mass spectroscopy analysis of the culture extracts was performed to confirm the presence of lipstatin in the context of *m/z* ratio. The confirmatory peak for the lipstatin observed at retention time of 11.33 min with a *m/z* of 491.369, represents the positive ionization $[M+H]^+$ of lipstatin (Fig. 4). Similar results were obtained by Sladič and group, (2014) during LC–MS analysis of lipstatin (Sladič et al. 2014). In another study, the *m/z* value for compound 6 was 492.371 indicating the positive ionization $[M+H]^+$ of lipstatin (Yang et al. 2019). It has been reported that the molecular weight of lipstatin is 491.7 and the exact mass of lipstatin is 491.361.

HPLC and LC–MS/MS analysis confirmed the lipstatin production potential of KD18 strain. The compound is utilised as an anti-obesity drug as it is an irreversible inhibitor of pancreatic lipase. Lipstatin is sold in the market under the trade name of Orlistat, the only FDA approved anti-obesity drug. The saturated form of lipstatin is tetrahydrolipstatin (THL) which is more stable as compared to the natural form of lipstatin. THL might be utilized for the treatment of giardiasis in future as it possesses the ability to inhibit the growth of *Giardia duodenalis* under in vitro conditions (Hahn et al. 2013).

Finally, the genome sequence might be utilized for genetic engineering to improve the production of secondary metabolites or alter the morphology of the bacteria. In addition, the genome sequence might aid in designing approaches to uncover the hidden metabolic pathways of cryptic clusters.
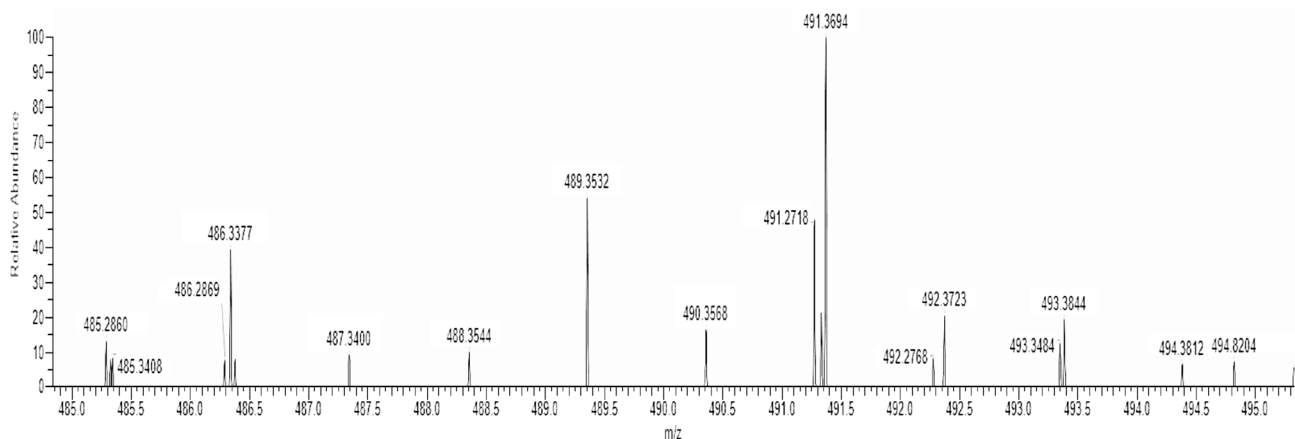
Declarations

Fig. 4 LC–MS/MS analysis of crude extract of *Streptomyces* sp. KD18 broth culture after 10 days of cultivation in production medium. MS chromatogram was recorded for m/z 491.361. MS spectrum of lipstatin with t_R (Retention time) = 11.33 min. Acquired *m/z* 491.369 corresponds to the positive ionization of lipstatin with theoretical *m/z* 491.361

References

Aziz RK, Bartels D, Best AA et al (2008) The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:1–15

Baltz RH (2019) Natural product drug discovery in the genomic era: realities, conjectures, misconceptions, and opportunities. J Ind Microbiol Biotechnol 46:281–299

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

Beld J, Sonnenschein EC, Vickery CR et al (2014) The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life. Nat Prod Rep 31:61–108

Benaud N, Edwards RJ, Amos TG et al (2021) Antarctic desert soil bacteria exhibit high novel natural product potential, evaluated through long-read genome sequencing and comparative genomics. Environ Microbiol 23:3646–3664

Bible KC, Ryder M (2016) Evolving molecularly targeted therapies for advanced-stage thyroid cancers. Nat Rev Clin Oncol 13:403–416

Boetzer M, Henkel C, Jansen HJ et al (2011) Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27:578–579

Boldogköi Z, Murvai J, Fodor I (1995) G and C accumulation at silent positions of codons produces additional ORFs. Trends Genet 11:125–126

Chaudhari NM, Gupta VK, Dutta C (2016) BPGA-an ultra-fast pangenome analysis pipeline. Sci Rep 6:1–10

Chen AY, Jemal A, Ward EM (2009) Increasing incidence of differentiated thyroid cancer in the United States, 1988–2005. Cancer 115:3801–3807

Choi YJ, Lee J-E, Ji HD et al (2021) Tunicamycin as a novel redifferentiation agent in radioiodine therapy for anaplastic thyroid cancer. Int J Mol Sci 22:1077

Claessen D, Rozen DE, Kuipers OP et al (2014) Bacterial solutions to multicellularity: a tale of biofilms, filaments and fruiting bodies. Nat Rev Microbiol 12:115–124

Galanie S, Entwistle D, Lalonde J (2020) Engineering biosynthetic enzymes for industrial natural product synthesis. Nat Prod Rep 37:1122–1143

Goodfellow M (2012) Phylum XXVI. Actinobacteria phyl. nov. In: Bergey's manual® of systematic bacteriology. Springer, pp 33–2028

Goris J, Konstantinidis KT, Klappenbach JA et al (2007) DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 57:81–91

Goto Y, Li B, Claesen J et al (2010) Discovery of unique lanthionine synthetases reveals new mechanistic and evolutionary insights. PLoS Biol 8:e1000339

Hahn J, Seeber F, Kolodziej H et al (2013) High sensitivity of Giardia duodenalis to tetrahydrolipstatin (orlistat) in vitro. PLoS One 8:e71597

Ikeda H, Ishikawa J, Hanamoto A et al (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. Nat Biotechnol 21:526–531

Kim J-N, Kim Y, Jeong Y et al (2015) Comparative genomics reveals the core and accessory genomes of Streptomyces species. J Microbiol Biotechnol 25:1599–1605

Komaki H, Sakurai K, Hosoyama A et al (2018) Diversity of nonribosomal peptide synthetase and polyketide synthase gene clusters among taxonomically close Streptomyces strains. Sci Rep 8:1–11

Konstantinidis KT, Tiedje JM (2005) Towards a genome-based taxonomy for prokaryotes. J Bacteriol 187:6258–6264

Kumar MS, Verma V, Soni S, Rao ASP (2012) Identification and process development for enhanced production of Lipstatin from *Streptomyces toxytricini* (Nrrl 15443) and further downstream processing. Adv Bio Tech 11:11

Labeda DP, Goodfellow M, Brown R et al (2012) Phylogenetic study of the species within the family *Streptomycetaceae*. Antonie Van Leeuwenhoek 101:73–104

Letunic I, Bork P (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296

Luo Y, Cobb RE, Zhao H (2014) Recent advances in natural product discovery. Curr Opin Biotechnol 30:230–237

Majer HM, Ehrlich RL, Ahmed A et al (2021) Whole genome sequencing of *Streptomyces actuosus* ISP-5337, *Streptomyces sioyaensis* B-5408, and *Actinospica acidiphila* B-2296 reveals secondary metabolomes with antibiotic potential. Biotechnol Reports 29:e00596

Medema MH, Blin K, Cimermancic P et al (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res 39:W339–W346

Mulzer M, Tiegs BJ, Wang Y et al (2014) Total synthesis of tetrahydrolipstatin and stereoisomers via a highly regio-and diastereoselective carbonylation of epoxyhomoallylic alcohols. J Am Chem Soc 136:10814–10820

Nikodinovic J, Barrow KD, Chuck JA (2003) High yield preparation of genomic DNA from Streptomyces. Biotechniques 35:932–936. https://doi.org/10.2144/03355bm05

Seemann T (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069

Seipke RF (2015) Strain-level diversity of secondary metabolism in *Streptomyces albus*. PLoS One 10:e0116457

Shiraishi T, Yoshida T, Nakata S et al (2005) Tunicamycin enhances tumor necrosis factor–related apoptosis-inducing ligand–induced apoptosis in human prostate cancer cells. Cancer Res 65:6364–6370

Sladič G, Urukalo M, Kirn M et al (2014) Identification of lipstatin-producing ability in *Streptomyces virginiae* CBS 314.55 using dereplication approach. Food Technol Biotechnol 52:276–284

Stalder H, Oesterhelt G, Borgström B (1992) Tetrahydrolipstatin: degradation products produced by human carboxyl-ester lipase. Helv Chim Acta 75:1593–1603

Telfer TJ, Richardson-Sanchez T, Gotsbacher MP et al (2019) Analogues of desferrioxamine B (DFOB) with new properties and new functions generated using precursor-directed biosynthesis. Biometals 32:395–408

Vicente CM, Thibessard A, Lorenzi J-N et al (2018) Comparative genomics among closely related *Streptomyces* strains revealed specialized metabolite biosynthetic gene cluster diversity. Antibiotics 7:86

Wang Y, Coleman-Derr D, Chen G, Gu YQ (2015) OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res 43:W78–W84

Williams GJ (2013) Engineering polyketide synthases and nonribosomal peptide synthetases. Curr Opin Struct Biol 23:603–612

Wu M, Scott AJ (2012) Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. Bioinformatics 28:1033–1034

Wyszynski FJ, Lee SS, Yabe T et al (2012) Biosynthesis of the tunicamycin antibiotics proceeds via unique exo-glycal intermediates. Nat Chem 4:539–546

Yang H, Zhang Z, Yan L et al (2019) Characterization of lipstatin and the minor components from *Streptomyces toxytricini* fermentation broth by HPLC–ESI–Q-TOF–MS. Chromatographia 82:1791–1800

Zerouki C, Bensalah F, Kuittinen S et al (2021) Whole-genome sequencing of two *Streptomyces* strains isolated from the sand dunes of Sahara. BMC Genomics 22:1–21. https://doi.org/10.1186/s12864-021-07866-x