SHORT COMMUNICATION

Draft genome sequence of *Streptomyces tunisialbus* DSM 105760[T]

Ameni Ayed[1] · Daniel Wibberg[2] · Imène Zendah el Euch[3] · Marcel Frese[3] · Ferid Limam[1] · Norbert Sewald[3]

Abstract

Streptomyces strains are well known as promising source of bioactive secondary metabolites, important in ecology, biotechnology and medicine. In this study, we present the draft genome of the new type strain *Streptomyces tunisialbus* DSM 105760[T] (=JCM 32165[T]), a rhizospheric bacterium with antimicrobial activity. The genome is 6,880,753 bp in size (average GC content, 71.85%) and encodes 5802 protein-coding genes. Preliminary analysis with antiSMASH 5.1.2. reveals 34 predicted gene clusters for the synthesis of potential secondary metabolites, which was compared with those of *Streptomyces varsoviensis* NRRL ISP-5346.

Keywords *Streptomyces tunisialbus* · Draft genome · AntiSMASH 5.1.2 · Secondary metabolites

Introduction

Since the discovery of streptomycin in 1944 from *Streptomyces griseus,* the first antibiotic isolated from bacteria (Schatz et al. 1944), Actinobacteria have obtained large attention from the scientific community, mainly due to their ability in producing a wide spectrum of bioactive compounds with antimicrobial and antitumor activities (Demain 2014; Barka et al. 2016). Nowadays, this class provided more than 65% of antibiotics and antifungals used in medicine; including over 10,000 bioactive compounds that were produced by the members of the genus *Streptomyces* (Bérdy 2005, 2012; De Lima et al. 2012). This genus, proposed for the first time by Waksman and Henrici (1943), belong to the Gram-positive bacteria and is ubiquitous in nature (Kämpfer 2012). The classification of *Streptomyces* species is generally carried out using a polyphasic taxonomic approach, including chemotaxonomic, phenotypic, and genotypic characteristics

(Anderson and Wellington 2001). Currently (Status April 2020), about 841 species of the genus *Streptomyces* with valid names have been described and published so far (https://www.bacterio.net/streptomyces.html). Thus, several studies have been conducted to find new strains isolated from extreme habitats (Tiwari and Gupta 2012; Goodfellow 2013; Zhang et al. 2016). To cope with the major health problem of antibiotic-resistant microorganisms, many laboratories have focused on the discovery of new bioactive molecules produced from new species belonging to this genus. In fact, whole-genome sequencing (wgs) have opened new potential for better exploitation of useful secondary metabolites produced by this genus (Harrison and Studholme 2014; Lee et al. 2018; Ward and Allenb 2018). This fundamental technology was supported by the use of new bioinformatics tools like antiSMASH, which adds several new features, including prediction of gene cluster (Blin et al. 2017).

In this study, the genome of a new strain *Streptomyces tunisialbus* DSM 105760[T], which produces a broad spectrum of secondary metabolites with antibacterial and antifungal effects (Ayed et al. 2018), was sequenced. The genome sequence of *S. tunisialbus* was established to broaden the genomic basis for functional and comparative analyses focusing on secondary metabolite biosynthesis clusters of the strain.

Materials and methods

The genomic DNA was extracted from a culture grown 3 days in ISP2 medium (Shirling and Gottlieb 1966), using cetyltrimethylammonium bromide (CTAB)-based method (Hopwood 2000). The quality of the DNA was assessed by gel-electrophoresis, and its quantity was determined by a fluorescence-based method using the Quant-iT PicoGreen dsDNA kit (Invitrogen) and the Tecan Infinite 200 Microplate Reader (Tecan Deutschland GmbH). For sequencing of the *S. tunisialbus* DSM 105760[T] genome, 4 µg of purified DNA was used to generate a paired library (TruSeq DNA PCR-Free Library Prep Kit, Illumina). The draft genome sequence of strain *S. tunisialbus* was established by sequencing on an Illumina MiSeq system. To assemble the obtained and processed reads, the GS de novo Assembler software version 2.8 (Roche, Mannheim, Germany) with default setting was applied as recently described (Yücel et al. 2017). Annotation of the draft genome was performed by applying the GenDB 2.0 system (Meyer et al. 2003) and Prokka version 1.11 (Seemann 2014). To identify clusters involved in secondary metabolite biosynthesis, the annotated genome was analyzed by applying antiSMASH 5.1.2 (Blin et al. 2017). Based on the 16S analysis in Ayed et al., *S. varsoviensis* NRRL ISP-5346 was determined as closest relative with a sequenced genome. *S. varsoviensis* NRRL ISP-5346 was found to produce multiple hygrolide sub-families, e.g., hygrobafilomycins (JBIR-100 and hygrobafilomycin) and bafilomycins (bafilomycins C1 and D) (Molloy et al. 2016). Finally, comparative genome analysis for *S. tunisialbus* DSM 105760[T] and *S. varsoviensis* NRRL ISP-5346 was performed by means of the comparative genomics tool EDGAR (Blom et al. 2009, 2016). To determine the relatedness between the *S. tunisialbus* DSM 105760[T] and *S. varsoviensis* NRRL ISP-5346 (Accession number: JOBF0100001-118), an average nucleotide identity analysis (ANI), average amino acid identity analysis (AAI) (Konstantinidis and Tiedje 2005) and a DDH estimation were performed as described previously (Meier-Kolthoff et al. 2013).

Results and discussion

The MiSeq sequencing run (2×300 bp) resulted in 2,092,356 reads yielding approx. 627 Mb sequence information for the sample representing on average a coverage of about 90-fold. The final draft genome established by the GS Assembler software (version 2.8, Roche) consists of 35 contigs (N50: 455,831 bp) and 20 scaffolds (N50: 678,417 bp). The annotation of the draft genome of *S. tunisialbus* showed a linear chromosome which has a size of 6,880,753 bp and encodes 5802 protein-coding sequences (CDS)s. The chromosome also contains 72 tRNA genes and eight rRNA operons (16S-23S-5S rRNA). The average GC content is 71.85% (Table 1).

An ANI value of 84.93%, an AAI value of 87.43% and a DDH value of 25.00% were calculated for *S. tunisialbus* DSM 105760[T] and *S. varsoviensis* NRRL ISP-5346. High ANI, AAI values (above 95%) and high DDH estimations (Formula 2 above 70%) were usually observed for bacterial isolates representing the same species and are considered to indicate bacterial species demarcation previously (Konstantinidis and Tiedje 2005). Therefore, both strains are closely related, but do not belong to the same species.

By applying EDGAR, a comparative analysis was performed for *S. tunisialbus* DSM 105760[T] and *S. varsoviensis* NRRL ISP-5346. It appeared that 3830 genes corresponding to 66% and 55% of all genes identified in both draft genomes represent the core set of genes shared by both isolates. These core genes mainly encode for housekeeping functions, e.g., glycolysis. However, *S. tunisialbus* DSM 105760[T] possesses 1972 singletons, whereas 3062 singletons were identified for *S. varsoviensis* NRRL ISP-5346. Most of the *S. tunisialbus* DSM 105760[T] singletons were annotated as 'hypothetical' genes or transposases. However, genes encoding enzymes potentially involved in the synthesis of secondary metabolites were identified within the unique genes of *S. tunisialbus* DSM 105760[T]. In addition, a few unique genes encoding for different cytochromes were found. For *S. varsoviensis* NRRL ISP-5346, the clusters found to produce hygrobafilomycins (JBIR-100 and hygrobafilomycin) and bafilomycins (bafilomycins C1 and D) were unique.

Table 1 General features of *S. tunisialbus* DSM 105760[T] and *S. varsoviensis* NRRL ISP-5346

Features	*S. tunisialbus* DSM 105760[T]	*S. varsoviensis* NRRL ISP-5346
Length (bp)	6,880,753	8,588,807
Status	Draft genome (35 contigs)	Draft genome (118 contigs)
G+C content (%)	71.85	72.40
CDS	5802	7229
rRNAs operons	8	NA
tRNAs genes	72	68

Table 2 Secondary metabolite cluster of *S. tunisialbus* DSM 105760[T]

Region	Type	Most similar known cluster	
Region 1	NRPS-like,lassopeptide	Citrulassin B	100.00%
Region 2	Terpene	2-Methylisoborneol	100.00%
Region 3	T2PKS, T1PKS, NRPS, terpene	JBIR-76/JBIR-77	68.00%
Region 4	NRPS	C-1027	7.00%
Region 5	Terpene	Calicheamicin	2.00%
Region 6	Ectoine	Ectoine	100.00%
Region 7	Lanthipeptide	SapB	75.00%
Region 8	Linaridin	Pentostatine/vidarabine	12.00%
Region 9	Melanin	Melanin	28.00%
Region 10	Melanin	Melanin	28.00%
Region 11	Terpene	Aristeromycin	6.00%
Region 12	Lassopeptide	Lagmysin	60.00%
Region 13	T3PKS	Herboxidiene	5.00%
Region 14	T3PKS, ladderane	Vazabitide A	28.00%
Region 15	Butyrolactone	Griseoviridin/fijimycin A	8.00%
Region 16	T1PKS	Elaiophylin	8.00%
Region 17	T1PKS	Guadinomine	7.00%
Region 18	T1PKS, thiopeptide, LAP, NRPS-like, terpene, NRPS	Aureothin	100.00%
Region 19	hglE-KS	a201a	5.00%
Region 20	Terpene	2-Methylisoborneol	75.00%
Region 21	hglE-KS, T3PKS	Alkylresorcinol	100.00%
Region 22	NRPS, T1PKS, terpene	Althiomycin	100.00%
Region 23	Bacteriocin, NRPS	Kirromycin	16.00%
Region 24	T1PKS, Terpene, bacteriocin	Nystatin A1	63.00%
Region 25	NRPS	Deimino-antipain	66.00%
Region 26	NRPS, terpene, T1PKS, transAT-PKS, PKS-like	Netropsin	100.00%
Region 27	Siderophore	Vibrioferrin	18.00%
Region 28	Terpene	Asukamycin	4.00%
Region 29	Terpene	Hopene	69.00%
Region 30	Indole	AT2433-A1	14.00%
Region 31	Butyrolactone		
Region 32	NRPS	Bacillibactin	46.00%
Region 33	T1PKS, NRPS, ladderane, arylpolyene, NRPS-like	Atratumycin	55.00%
Region 34	siderophore		

To identify all clusters involved in secondary metabolite biosynthesis, the *S. tunisialbus* DSM 105760[T] genome was analyzed by means of antiSMASH 5.1.2. The platform revealed the presence of 34 predicted gene clusters encode for potential secondary metabolites, such as antibiotics, antitumor compounds, fungicide, and lantibiotics. Likewise, six of these gene clusters are terpenes, such 2-methylisoborneol [100%] a volatile organic compound well known for its biological activity, two for melanin biosynthesis and one for Ectoine (Table 2). *S. varsoviensis* NRRL ISP-5346 has a similar portfolio of secondary metabolite clusters (Table 3).

The presence of a large number of cryptic secondary metabolite gene clusters in its genome, offers unsuspected potential for synthesizing bioactive compounds. In addition, our study shows that sequencing of new strains allows the discovery of new secondary metabolites relevant in ecology, biotechnology and medicine.

Table 3 Secondary metabolite cluster of *S. varsoviensis* NRRL ISP-5346

Region	Type	Most similar known cluster	
Region 1	Terpene, T1PKS, NRPS-like, phosphonate, nucleoside, lanthipeptide	Phosphinothricintripeptide	17.00%
Region 2	NRPS-like, T1PKS	Miharamycin A/miharamycin B	40.00%
Region 3	PKS-like	A-201A	12.00%
Region 4	Terpene		
Region 5	NRPS	Isocomplestatin	50.00%
Region 6	Lanthipeptide	AmfS	80.00%
Region 7	Lassopeptide	Anantin B1/anantin B2	40.00%
Region 8	Furan, NRPS	Ficellomycin	3.00%
Region 9	NRPS-like, NRPS	Friulimicin A/friulimicin B/friulimicin C/friulimicin D	27.00%
Region 10	Amglyccycl		
Region 11	NRPS	Streptobactin	47.00%
Region 12	Amglyccycl	Hygromycin A	51.00%
Region 13	Fused, NRPS-like, TfuA-related, NRPS, T1PKS	Thiovarsolin A/thiovarsolin B/thiovarsolin C/thiovarsolin D	83.00%
Region 14	NRPS	rimosamide	42.00%
Region 15	T2PKS,PKS-like, oligosaccharide, LAP	Dutomycin	36.00%
Region 16	Bacteriocin		
Region 17	Terpene		
Region 18	hglE-KS, T1PKS	Tetronasin	9.00%
Region 19	Lanthipeptide		
Region 20	Ectoine	Ectoine	100.00%
Region 21	NRPS, T2PKS	Kosinostatin	6.00%
Region 22	LAP, thiopeptide	Jomthonic acid A/jomthonic acid B/jomthonic acid C	5.00%
Region 23	Thiopeptide, LAP		
Region 24	Terpene	Ansamitocin P-3	4.00%
Region 25	NRPS,CDPS	Purincyclamide	100.00%
Region 26	T1PKS	Chlorizidine A	57.00%
Region 27	Lanthipeptide		
Region 28	NRPS, T3PKS	Feglymycin	42.00%
Region 29	T1PKS, NRPS, aminocoumarin	Bafilomycin B1	83.00%
Region 30	CDPS		
Region 31	Terpene	Ebelactone	5.00%
Region 32	Siderophore		
Region 33	Siderophore	Ficellomycin	3.00%
Region 34	Siderophore		
Region 35	Terpene	Hopene	61.00%
Region 36	Bacteriocin		
Region 37	NRPS		

References

Anderson AS, Wellington EMH (2001) The taxonomy of *Streptomyces* and related genera. Int J Syst Evol Microbiol 51:797–814

Ayed A, Slama N, Mankai H, Bachkouel S, ElKahoui S, Tabbene O, Limam F (2018) *Streptomyces tunisialbus* sp. nov., a novel *Streptomyces* species with antimicrobial activity. Antonie Van Leeuwenhoek 111:1571–1581

Barka EA, Vatsa P, Sanchez L, Gaveau-Vaillant N, Jacquard C, Klenk HP (2016) Taxonomy, physiology, and natural products of Actinobacteria. Microbiol Mol Biol Rev 80:1–43

Bérdy J (2005) Bioactive microbial metabolites. J Antibiot 58:1–26

Bérdy J (2012) Thoughts and facts about antibiotics; where we are now and where we are heading. J Antibiot 65:385–395

Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de Los Santos ELC, Kim HU, Nave M, Dickschat JS, Mitchell DA, Shelest E, Breitling R, Takano E, Lee SY, Weber T, Medema MH (2017) antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. Nucleic Acids Res 45(W1):36–41

Blom J, Albaum SP, Doppmeier D, Pühler A, Vorhölter FJ, Zakrzewski M, Goesmann A (2009) EDGAR: a software framework for the comparative analysis of prokaryotic genomes. BMC Bioinform 10:154–159

Blom J, Kreis J, Spänig S, Juhre T, Bertelli C, Ernst C, Goesmann A (2016) EDGAR 2.0: an enhanced software platform for comparative gene content analyses. Nucleic Acids Res 44:22–28

De Lima PRE, Da Silva IR, Martins MK, De Azevedo JL, de Araújo JM (2012) Antibiotics produced by *Streptomyces*. Braz J Inf Dis 16(5):466–471

Demain A (2014) Importance of microbial natural products and the need to revitalize their discovery. J Ind Microbiol Biotechnol 41:185–201

Goodfellow M (2013) Actinobacterial diversity as a source of new drugs. Microbiologist 14:8–12

Harrison J, Studholme DJ (2014) Recently published *Streptomyces* genome sequences. Microb Biotechnol 7:373–380

Hopwood DA (2000) Practical *Streptomyces* genetics, 518. Norwich, England, pp 170–171

Kämpfer P (2012) Genus I. *Streptomyces* Waksman and Henrici 1943, 339 emend. Witt and Stackebrandt 1990, 370, emend. Wellington, Stackebrandt, Sanders, Wolstrup and Jorgensen, 1992, 159. In: Goodfellow M, Kämpfer P, Busse H-J, Trujillo M, Suzuki KE, Ludwig W, Whitman WB (eds) Bergey's Manual of Systematic Bacteriology, vol 5, 2nd edn. Springer, New York, pp 1455–1767

Konstantinidis KT, Tiedje JM (2005) Genomic insights that advance the species definition for prokaryotes. Proc Natl Acad Sci USA 102:2567–2572

Lee LH, Chan KG, Stach J, Elizabeth Wellington EMH, Goh BH (2018) Editorial: the search for biological active agent(s) From actinobacteria. Front Microbiol 9:824–829

Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinform 14:60

Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, Clausen J, Kalinowski J, Linke B, Rupp O, Giegerich R, Pühler A (2003) GenDB—an open source genome annotation system for prokaryote genomes. Nucleic Acids Res 31:2187–2195

Molloy EM, Tietz JI, Blair PM, Mitchell DA (2016) Biological characterization of the hygrobafilomycin antibiotic JBIR-100 and bioinformatic insights into the hygrolide family of natural products. Bioorg Med Chem 24:6276–6290

Schatz A, Bugle E, Waksman SA (1944) Streptomycin, a substance exhibiting antibiotic activity against Gram-positive and Gram-negative bacteria. Proc Soc Exp Biol Med 55:66–69

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

Shirling EB, Gottlieb D (1966) Methods for characterization of *Streptomyces* species. Int J Syst Bacteriol 16:313–340

Tiwari K, Gupta RK (2012) Rare actinomycetes: a potential storehouse for novel antibiotics. Crit Rev Biotechnol 32:108–132

Waksman SA, Henrici AT (1943) The nomenclature and classification of the actinomycetes. J Bacteriol 46:337–341

Ward AC, Allenby NEE (2018) Genome mining for the search and discovery of bioactive compounds: the *Streptomyces* paradigm. FEMS Microbiol Lett. https://doi.org/10.1093/femsle/fny240

Yücel O, Wibberg D, Philipp B, Kalinowski J (2017) Genome sequence of the bile salt-degrading bacterium *Novosphingobium* sp. strain Chol11, a model organism for bacterial steroid catabolism. Genome Announc 6(1):e01372–e1417

Zhang L, Ruan C, Peng F, Deng Z, Hong K (2016) *Streptomyces arcticus* sp. nov., isolated from frozen soil. Int J Syst Evol Microbiol 66:1482–1487