Progress in Microbes and Molecular Biology

HH PUBLISHER

Genome Report

Whole-Genome Sequence of *Streptomyces pluripotens* strain MUM 16J, a Potential Resource of Glycopeptide Antibiotic and Biocontrol Agent against Biofilm-forming Bacteria

Priyia Pusparajah[1], Jodi Woan-Fei Law[2,3], Kok-Gan Chan[3,4,5], Bey-Hing Goh[6,7], Kar-Wai Hong[3*], Learn-Han Lee[2,3,8*], Loh Teng-Hern Tan[3,8*]

Article History

Received: 28 March 2023;

Received in Revised Form: 20 April 2023;

Accepted: 29 April 2023;

Available Online: 02 May 2023

[1]Medical Health and Translational Research Group (MHTR), Jeffrey Cheah School of Medicine and Health Sciences, Monash University Malaysia, 47500 Bandar Sunway, Selangor Darul Ehsan, Malaysia; priyia.pusparajah@monash.edu (PP)

[2]Next-Generation Precision Medicine and Therapeutics Research Group (NMeT), Jeffrey Cheah School of Medicine and Health Sciences, Monash University Malaysia, Bandar Sunway 47500, Selangor Darul Ehsan, Malaysia; Jodi.Law1@monash.edu (JW-FL)

[3]Novel Bacteria and Drug Discovery Research Group (NBDD), Microbiome and Bioresource Research Strength (MBRS), Jeffrey Cheah School of Medicine and Health Sciences, Monash University Malaysia, Bandar Sunway 47500, Selangor Darul Ehsan, Malaysia

[4]International Genome Centre, Jiangsu University, Zhenjiang, China.

[5]Institute of Biological Sciences, Faculty of Science, University of Malaya, 50603, Kuala Lumpur, Malaysia; kokgan@um.edu.my (KGC)

[6]Biofunctional Molecule Exploratory Research Group (BMEX), School of Pharmacy, Monash University Malaysia, 47500 Bandar Sunway, Selangor Darul Ehsan, Malaysia; goh.bey.hing@monash.edu (B-HG)

[7]College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, Zhejiang, People's Republic of China

[8]Innovative Bioprospection Development Research Group (InBioD), Clinical School Johor Bahru, Jeffrey Cheah School of Medicine and Health Sciences, Monash University Malaysia, Johor Bahru 80100, Malaysia

*Corresponding author: Kar-Wai Hong, Learn-Han Lee; Novel Bacteria and Drug Discovery Research Group (NBDD), Microbiome and Bioresource Research Strength (MBRS), Jeffrey Cheah School of Medicine and Health Sciences, Monash University Malaysia, 47500, Selangor, Malaysia; hong.karwai@monash.edu (KWH), lee.learn.han@monash.edu (L-HL); Loh Teng-Hern Tan; Innovative Bioprospection Development Research Group (InBioD), Clinical School Johor Bahru, Jeffrey Cheah School of Medicine and Health Sciences, Monash University Malaysia, Johor Bahru 80100, Malaysia; loh.teng.hern@monash.edu (LT-HT)

Abstract: *Streptomyces pluripotens* strain MUM 16J is a Gram-positive filamentous bacteria isolated from mangrove soil in Kampung Trombol area of Kuching, Sarawak. The draft genome of *Streptomyces pluripotens* strain MUM 16J consists of 7,671,699 bp assembled into 164 contigs. The GC content of the genome is 70.06 %, and the sequencing coverage of 108×. Its genome consists of 6,781 predicted genes, 6,459 protein-coding genes, and 79 RNA-coding genes (tRNA: 71, rRNA: 8). Here we report the draft genome of this strain to highlight its potential in producing glycopeptide antibiotic as well as catechol, a biocontrol agent of biofilm-producing bacteria.

Keywords: *Streptomyces pluripotens*; genome; antibiofilm; glycopeptide; catechol; Malaysia

1. Introduction

Streptomyces are traditionally considered as Gram-positive spores-producing, soil-dwelling and mycelium-forming filamentous bacteria with enormous potential for biotechnological and pharmaceutical applications [1-4]. *Streptomyces* has been a rich source of valuable medicinal chemicals, ranging from antibiotic to antifungal, antioxidant, antiparasitic to even anticancer compounds [5-12]. Other than medical and pharmaceutical industry, *Streptomyces* also play an important role in various industrial bioprocesses such as industrial fermentative production [13-18]. The study of microbes in pristine environments such as the unexplored regions of rainforest, peat swamp forest and mangrove forest represent a promising avenue of research for the discovery of new biologically active compounds [19-23]. In this study, we present the genomic features of *S. pluripotens* strain MUM 16J derived from an unexplored mangrove environment and highlight its capability to produce glycopeptide antibiotic and catechol (biocontrol agent of biofilm-forming bacteria).

2. Data description

Streptomyces sp. strain MUM 16J was isolated from the mangrove soil in Kampung Trombol area of Kuching, Sarawak. Bacteria were routinely grown in International *Streptomyces* Project (ISP) 2 medium at 28°C [24]. The genomic DNA was extracted using the MasterPure Gram-positive DNA purification kit (Lucigen, Middleton, WI, USA), according to the recommended procedures of the manufacturer [25]. Subsequently, the extracted DNA was subjected to RNase treatment. Prior to sequencing library construction, the quantity and quality of the extracted genomic DNA were measured using a Qubit 2.0 fluorometer and a NanoDrop 2000 spectrophotometer (both Thermo Fisher Scientific, Waltham, MA, USA), respectively [26]. The sequencing library was constructed using Nextera™ DNA Sample Preparation kit (Nextera, USA) [27]. The library profile, size distribution and concentration of the sequencing library were evaluated using Bioanalyser 2100 High Sensitivity DNA kit (Agilent Technologies, Palo Alto, CA) prior to performing

paired-end sequencing on MiSeq platform with MiSeq Reagent Kit 2 (2×250bp; Illumina Inc., Madison, WI, USA) [28, 29].

Upon completion of the whole genome sequencing, the quality of the sequencing raw reads was evaluated using FastQC (version 0.11.9) [30]. Subsequently, sequencing reads with a mapping quality below Q20 were removed from subsequent analysis using Trimmomatic (version 0.39) [31]. The quality-trimmed reads were assembled *de novo* using SPAdes (version 3.14.1) [32]. The assembled genome was evaluated based on contiguity and completeness with single-copy orthologs using the QUAST (version 5.2.0) [33] and Benchmarking Universal Single-Copy Orthologs (BUSCO) (version 5.4.4) [34], respectively. The lineage datasets in the BUSCO analysis were streptomycetales_odb10. The assembled sequences were annotated using Prokka (version 1.14.6) and NCBI Prokaryotic Genome Annotation Pipeline (version 4.13). The genomic features of *S. pluripotens* strain MUM 16J are shown in Table 1. From the BUSCO analysis, a total of 1,579 BUSCO markers were searched, of which there are 1,571 complete and single-copy BUSCO markers identified, and 6 complete and duplicated BUSCO markers found. No fragmented BUSCO markers were found and 2 BUSCO markers were missing.

Table 1. Genomic features of *Streptomyces pluripotens* strain MUM 16J.

	Streptomyces pluripotens MUM 16J
Strain	MUM 16J
Total number of contigs	164
Genome size (bp)	7,671,699
Sequencing coverage	108×
GC (%)	70.06
N50 (bp)	164,408
N90 (bp)	40,084
L50 (bp)	15
L90 (bp)	47
Total number of predicted genes	6,781
Total number of protein-coding genes	6,459
Total number of RNA-coding genes	79 (tRNA-coding genes: 71, rRNA-coding genes: 8)
Total number of ncRNA-coding genes	3
Total number of pseudogenes	243
BioSample	SAMN16860772
BioProject	PRJNA679886
GenBank Assembly Accession Number	GCA_022414615.1
GenBank Nucleotide Accession Number	JADWYN000000000.1

From the assembled genome, the 16S ribosomal RNA (rRNA) gene of *Streptomyces pluripotens* strain MUM 16J was predicted using Barrnap (version 0.9) [35], and the predicted 16S rRNA gene sequence was searched against EzBioCloud Database [36, 37], a well-curated database of 16S rRNA sequences and bacterial genomes. From the 16S gene analysis using the EzBioCloud Database, the 16S rRNA gene of strain MUM 16J has a similarity and completeness of 100%, respectively, with *S. pluripotens* strain MUSC 135, suggesting strain MUM 16J is very likely to be *S. pluripotens*.

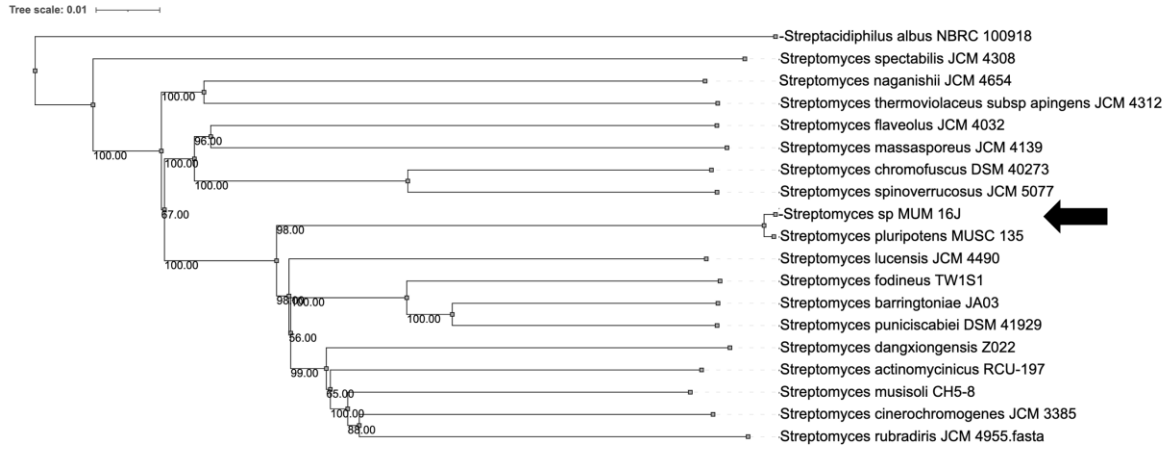

Figure 1. Position of *Streptomyces* sp. strain MUM 16J (arrow) in phylogenetic tree comprising selecting species of *Streptomyces*, in which the phylogenetic tree was generated using the GGDC distance matrix.

The identities of the strain MUM 16J were also determined by a whole genome-based taxonomic analysis via Type (Strain) Genome Server (TYGS) [38], and digital DNA-DNA hybridization (dDDH) calculations via Genome-to-Genome Distance Calculator (GGDC) (version 3.0) [39-41]. The digital DDH values were calculated using GGDC distance formula d_4, which is the sum of all identities found in the high-score segment pairs (HSPs) divided by the total length of all HSPs. Supplementary Table S1 shows the analysis findings of GGDC. Figure 1 shows the position of strain MUM 16J in the phylogenetic tree comprising selecting species of *Streptomyces*, in which the phylogenetic tree was generated using the GGDC distance matrix (Kendall-Colijn test, $P < 0.05$).

The genome of *Streptomyces pluripotens* strain MUM 16J was used as a query to compare against the reference strains, namely *S. pluripotens* strains MUSC 135 (NCBI Accession number: CP021080.1) and strain MUSC 137 (NCBI Accession number: CP022433.1), using BLAST [42]. This is to identify the accessory genomes or non-core genome of strain MUM16J [43]. The accessory genome of strain MUM16J was analyzed using

antiSMASH (version 7.0) [44-46] and SeMPI (version 2.0) [47, 48] in order to identify its biosynthetic gene clusters. From the analysis output of antiSMASH, a glycopeptide antibiotic-producing BGS was identified on NZ_JADWYN010000070.1. From the analysis output of SeMPI, two contigs, namely NZ_JADWYN010000151.1 and NZ_JADWYN010000152.1, were predicted to carry the gene which is responsible for the biosynthesis of catechol, in which catechol is a biofilm-inhibiting compound [49, 50].

The draft genome sequence of *S. pluripotens* strain MUM 16J has been deposited at DDBJ/EMBL/GenBank under the accession number JADWYN010000000.1. The version described in this paper is the first version. The data are publicly available at NCBI GenBank under the BioProject accession number PRJNA679886, and the BioSample accession number SAMN16860772.

Author Contributions: Writing—original draft preparation, PP; conceptualization, K-WH and B-HG, L-HL, LT-HT; methodology and data analysis, PP, LT-TH, JW-FL and K-WH; validation, LT-HT, B-HG, L-HL and K-GC; review and editing, K-WH, LT-HT; resources, L-HL, and K-GC.

Funding: This work is supported by the Jeffrey Cheah School of Medicine and Health Sciences Strategic Grant Grant 2022 (Vote Number: STG-000107), awarded to PP.

Acknowledgments: The authors thank Hooi-Leng Ser and Vengadesh Letchumanan for their kind support and contribution to the project.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Law JWF, Letchumanan V, Tan LTH, *et al*. The rising of "Modern Actinobacteria" era. Prog Microbes Mol Biol 2020; 3(1).
2. Law JWF, Pusparajah P, Ab Mutalib NS, *et al*. A review on mangrove Actinobacterial diversity: The roles of Streptomyces and novel species discovery. Prog Microbes Mol Biol 2019; 2(1).
3. Mutalib NSA, Wong SH, Ser HL, *et al*. Bioprospecting of microbes for valuable compounds to mankind. Prog Microbes Mol Biol 2020; 3(1).
4. Lee L-H, Chan K-G, Stach J, *et al*. The search for biological active agent (s) from actinobacteria. Front Microbiol 2018; 9: 824.
5. Tan LTH, Lee LH, and Goh BH. The bioprospecting of anti-Vibrio Streptomyces species: Prevalence and applications. Prog Microbes Mol Biol 2019; 2(1).
6. Elsalami RM, Goh KW, Mahadi M, *et al*. The antibacterial activities of secondary metabolites derived from Streptomyces sp. Prog Microbes Mol Biol 2022; 5(1).
7. Kemung HM, Tan LTH, Chan KG, *et al*. Streptomyces sp. strain MUSC 5 from mangrove forest in Malaysia: Identification, antioxidant potential and chemical profiling of its methanolic extract. Prog Microbes Mol Biol 2020; 3(1).
8. Kemung HM, Tan LT-H, Chan K-G, *et al*. Antioxidant activities of Streptomyces sp. strain MUSC 14 from mangrove forest soil in Malaysia. BioMed Res Int 2020; 2020.
9. Mangzira Kemung H, Tan LT-H, Chan K-G, *et al*. Streptomyces sp. strain MUSC 125 from mangrove soil in Malaysia with anti-MRSA, anti-biofilm and antioxidant activities. Molecules 2020; 25(15): 3545.

10. Law JW-F, Ser H-L, Ab Mutalib N-S, *et al*. Streptomyces monashensis sp. nov., a novel mangrove soil actinobacterium from East Malaysia with antioxidative potential. Sci Rep 2019; 9(1): 3056.
11. Tan LT-H, Chan K-G, Chan CK, *et al*. Antioxidative potential of a Streptomyces sp. MUM292 isolated from mangrove soil. BioMed Res Int 2018; 2018.
12. Tan LTH, Mahendra CK, Yow YY, *et al*. Streptomyces sp. MUM273b: A mangrove-derived potential source for antioxidant and UVB radiation protectants. MicrobiologyOpen 2019; 8(10): e859.
13. Latha S, Sivaranjani G, and Dhanasekaran D. Response surface methodology: A non-conventional statistical tool to maximize the throughput of Streptomyces species biomass and their bioactive metabolites. Crit Rev Microbiol 2017; 43(5): 567-582.
14. Sanchez J, Yague P, and Manteca A. New insights in Streptomyces fermentations. Ferment Technol 2012; 1(2).
15. Tan LTH, Lee LH, and Goh BH. Critical review of fermentation and extraction of anti-Vibrio compounds from Streptomyces. Prog Microbes Mol Biol 2020; 3(1).
16. Manteca A, Alvarez R, Salazar N, *et al*. Mycelium differentiation and antibiotic production in submerged cultures of Streptomyces coelicolor. Appl Environ Microbiol 2008; 74(12): 3877-3886.
17. Barbuto Ferraiuolo S, Cammarota M, Schiraldi C, *et al*. Streptomycetes as platform for biotechnological production processes of drugs. Appl Microbiol Biotechnol 2021; 105(2): 551-568.
18. Chater KF. Recent advances in understanding Streptomyces. F1000Res 2016; 5: 2795.
19. Ong KS, Letchumanan V, Law JWF, *et al*. Microbes from peat swamp forest — The hidden reservoir for secondary metabolites? Prog Microbes Mol Biol 2020; 3(1).
20. Ser HL, Law JWF, Tan WS, *et al*. Whole genome sequence of Streptomyces colonosanans strain MUSC 93JT isolated from mangrove forest in Malaysia. Prog Microbes Mol Biol 2020; 3(1).
21. Ser HL, Tan LTH, Tan WS, *et al*. Whole-genome sequence of bioactive streptomycete derived from mangrove forest in Malaysia, Streptomyces sp. MUSC 14. Prog Microbes Mol Biol 2021; 4(1).
22. Ser HL, Tan WS, Yin WF, *et al*. Whole genome sequence of MUM116, a Bacillus species isolated from intertidal soil. Prog Microbes Mol Biol 2020; 3(1).
23. Lee L-H, Zainal N, Azman A-S, *et al*. Mumia flava gen. nov., sp. nov., an actinobacterium of the family Nocardioidaceae. Int J Syst Evol Microbiol 2014; 64(Pt_5): 1461-1467.
24. Law JWF, Tan LT-H, Letchumanan V, *et al*. Streptomyces griseiviridis sp. nov., A novel "Modern Actinobacteria" isolated from Malaysia mangrove soil. Prog Microbes Mol Biol 2023; 6(1).
25. Chan X-Y, Chen J-W, Adrian T-G-S, *et al*. Whole-genome sequence and fosfomycin resistance of Bacillus sp. strain G3(2015) isolated from seawater off the coast of Malaysia. Genome Announc 2017; 5(13): e00067-17.
26. Hong K-W, Koh C-L, Sam C-K, *et al*. Whole-genome sequence of N-acylhomoserine lactone-synthesizing and -degrading Acinetobacter sp strain GG2. J Bacteriol 2012; 194(22): 6318.
27. Chan K-G, Chong T-M, Adrian T-G-S, *et al*. Whole-genome sequence of Stenotrophomonas maltophilia ZBG7B reveals its biotechnological potential. Genome Announc 2015; 3(6): e01442-15.
28. Kher H-L, Krishnan T, Letchumanan V, *et al*. Characterization of quorum sensing genes and N-acyl homoserine lactones in Citrobacter amalonaticus strain YG6. Gene 2019; 684: 58-69.
29. Goh YX, Chan KG, and Hong KW. Whole-genome sequence of Chelatococcus daeguensis strain M38T9, Isolated from Ulu Slim hot spring in Malaysia. Prog Microbes Mol Biol 2022; 5(1).
30. Andrews S. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom 2010.

31. Bolger AM, Lohse M, and Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014; 30(15): 2114-2120.
32. Prjibelski A, Antipov D, Meleshko D, *et al*. Using SPAdes De Novo Assembler. Curr Protoc Bioinform 2020; 70(1): e102.
33. Gurevich A, Saveliev V, Vyahhi N, *et al*. QUAST: quality assessment tool for genome assemblies. Bioinformatics 2013; 29(8): 1072-1075.
34. Simão FA, Waterhouse RM, Ioannidis P, *et al*. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 2015; 31(19): 3210-3212.
35. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2014; 30(14): 2068-2069.
36. Chun J, Lee J-H, Jung Y, *et al*. EzTaxon: a web-based tool for the identification of prokaryotes based on 16S ribosomal RNA gene sequences. Int J Syst Evol Microbiol 2007; 57(10): 2259-2261.
37. Tan LT-H, Ser H-L, Yin W-F, *et al*. Investigation of antioxidative and anticancer potentials of Streptomyces sp. MUM256 isolated from Malaysia mangrove soil. Front Microbiol 2015; 6: 1316.
38. Meier-Kolthoff JP and Göker M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Comm 2019; 10(1): 2182.
39. Meier-Kolthoff JP, Auch AF, Klenk H-P, *et al*. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 2013; 14(1): 60.
40. Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, *et al*. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. Nucleic Acids Res 2021; 50(D1): D801-D807.
41. Meier-Kolthoff JP, Klenk H-P, and Göker M. Taxonomic use of DNA G+ C content and DNA–DNA hybridization in the genomic age. Int J Syst Evol Microbiol 2014; 64(Pt_2): 352-356.
42. Camacho C, Coulouris G, Avagyan V, *et al*. BLAST+: architecture and applications. BMC Bioinformatics 2009; 10(1): 421.
43. Lapierre P and Gogarten JP. Estimating the size of the bacterial pan-genome. Trends Genet 2009; 25(3): 107-110.
44. Blin K, Shaw S, Kloosterman AM, *et al*. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res 2021; 49(W1): W29-W35.
45. Blin K, Shaw S, Steinke K, *et al*. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 2019; 47(W1): W81-W87.
46. Blin K, Wolf T, Chevrette MG, *et al*. antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. Nucleic Acids Res 2017; 45(W1): W36-W41.
47. Zierep PF, Ceci AT, Dobrusin I, *et al*. SeMPI 2.0—A web server for PKS and NRPS predictions combined with metabolite screening in natural product databases. Metabolites 2021; 11(1): 13.
48. Zierep PF, Padilla N, Yonchev DG, *et al*. SeMPI: a genome-based secondary metabolite prediction and identification web server. Nucleic Acids Res 2017; 45(W1): W64-W71.
49. Razaviamri S, Wang K, Liu B, *et al*. Catechol-based antimicrobial polymers. Molecules 2021; 26(3): 559.
50. Kemung HM, Tan LTH, Khaw KY, *et al*. An optimized anti-adherence and anti-biofilm assay: Case study of zinc oxide nanoparticles versus MRSA biofilm. Prog Microbes Mol Biol 2020; 3(1).