Draft Genome Sequences of Two Polycyclic Tetramate Macrolactam Producers, *Streptomyces* spp. strains JV180 and SP18CM02

Yunci Qi[a*], Keshav K. Nepal[a*], Jennifer Greif[a], Cole Martini[a], Chad Tomlinson[b], Christopher Markovic[b], Catrina Fronick[b], Joshua A. V. Blodgett[a#]

[a]Department of Biology, Washington University in St. Louis, Missouri, USA

[b] McDonnell Genome Institute, Washington University School of Medicine, Missouri, USA

[*]Yunci Qi and Keshav Nepal contributed equally to this work. Author order was determined in order of increasing seniority.

[#]Corresponding Author: Washington University in St Louis, Department of Biology. 1 Brookings Dr., St Louis MO 63122. jblodgett@wustl.edu

ABSTRACT Here we report the draft genome sequences of two related *Streptomyces* spp. strains, JV180 and SP18CM02. Despite their respective isolation from Connecticut and Missouri (USA) soils, they are strikingly similar in gene content. Both belong to the *Streptomyces griseus* clade and harbor several secondary metabolite biosynthetic gene clusters (BGCs).

Streptomyces produce many diverse secondary metabolites that are medicinally relevant **(1)**. Recent advancements in genome sequencing have revealed *Streptomyces* as a reservoir of diverse BGCs. A high proportion of *Streptomyces* harbor BGCs encoding polycyclic tetramate macrolactams (PTMs), molecules of interest due to their structural complexity, atypical biosyntheses, and biomedical significance **(2-4)**. *Streptomyces* sp. strain JV180 and *Streptomyces* sp. strain SP18CM02 are soil isolates found to carry PTM BGCs. The isolation of strain JV180 was previously reported (**3),** while strain SP18CM02 was recently isolated using identical methods from calcium carbonate -treated soil collected at Tyson Research Center in Eureka, MO, using LTY medium supplemented with nalidixic acid.

Both strains were grown in Trypticase Soy Broth supplemented with 0.6 % glycine at 28 ˚C and genomic DNA was extracted with the Qiagen Dneasy UltraClear Microbial Kit for Illumina sequencing or phenol-chloroform **(5)** for PacBio sequencing. For PacBio sequencing, genomic

DNA was sheared to approximately 15 kb using Covaris g-tubes. PacBio SMRTbell Express Template Prep Kit 2.0, Barcoded Overhang Adapter Kit 8A, and Barcoded Overhang Adapter Kit 8B were used for library construction. Final library pools were assessed on an Agilent 2100 Bio-analyzer using a DNA 12000 kit. DNA polymerase binding complex was processed using PacBio Sequel Binding Kit 3.0 and sequencing primer v4. PacBio Sequel Sequencing Plate 3.0 and SMRT Cell 1M v3 Tray were used for the sequencing run. The resulting reads were evaluated by mean read length, which was 27,869 bp. PacBio read error correction, adapter trimming, and assembly were done in HGAP4. Default parameters were used with the exception of the advanced parameter: "Aggressive Mode". The assembly was polished in Pilon v.1.23 with Illumina reads, which were obtained as described below **(6-7).**

For Illumina sequencing, dual indexed libraries were constructed with 2 µg of gDNA utilizing the TruSeq PCR-free Library Prep Kit (Illumina) on the SciClone NGS instrument (Perkin Elmer) targeting 550bp inserts. 50 µL of gDNA was fragmented on a LE200 Covaris instrument and the library was evaluated on a LabChip GX (Perkin Elmer). The library was sequenced on a NovaSeq 6000 (Illumina) S4 300 cycle flow cell for 2 x 150 paired-end reads, and NovaSeq Real Time Analysis v. 3.3.3 software (Illumina) was used to ensure ≥80% of resulting bases were ≥Q30. Flexbar v. 3.4 **(8)** was used for adapter trimming and MEGAHIT **(9)** was used for genome assembly with default parameters. Genomic features were annotated by Rapid Annotations using Subsystems Technology version 2.0 **(10)** and AntiSMASH **(11)** (Table 1).

Table 1. Genomic features of *Streptomyces* sp. strains JV180 and SP18CM02

Streptomyces sp. strain	Sequencing method	Reads	Sequenced (bp)	Genome size (bp)	Fold coverage	Contigs	N50	G+C (%)	CDS	RNAs
JV180	PacBio	920,865	2,597,107,056	8,006,517	324.37	5	7,244,875	72.6	7353	83
	Illumina	22,533,770	3,402,599,270							
SP18CM02	Illumina	16,938,420	2,557,701,420	8,024,422	318.74	344	431,047	72.7	7374	76

Strains JV180 and SP18CM02 share 100% identical 16S rRNA sequences, and their closest match in Genbank was *S. californicus* strain NRRL B-3320 (99.93%), a strain that has a similar genome size (8,023,053 bp) which belongs to the *Streptomyces griseus* clade **(12)**. MUSCLE pairwise alignments **(13)** revealed that strains JV180 and SP18CM02 share 99.93, 99.76, 99.29, 99.91, and

99.53 % nucleotide identities for *atp*D, *gyr*B, *rec*A, *rpoB*, and *trp*B, respectively. Strains JV180 and SP18CM02 possess 34 and 35 BGCs, respectively, and all strain JV180 BGCs are conserved in strain SP18CM02.

Accession number(s). The draft genome sequences of strain JV180 and SP18CM02 were deposited in DDBJ/ENA/GeneBank under accession numbers JACGMP000000000 and JACGMQ000000000, respectively. SRA accession numbers for strain JV180 are SRR12430886 and SRR12430887 for Illumina and PacBio sequencing respectively, and SRR12430885 for strain SP18CM02. The version described in this paper is the first version.

ACKNOWLEDGEMENTS: This project was funded through a McDonnell Genome Institute 2018 Symposium Pilot Project funded by the Office of the Dean of the School of Medicine- Washington University Institute of Clinical and Translational Sciences (ICTS) and Illumina, Inc (San Diego CA) awarded to J Blodgett. Additional project support was made available through the McDonnell Genome Institute 2018 Symposium Pilot Project Fund at Washington University as well as the NIH/National Center for Advancing Translational Sciences (NCATS) grant UL1TR002345. This material is also based upon work supported by the National Science Foundation under NSF-CAREER 1846005 to J Blodgett.

REFERENCES

1. Nepal KK, Wang G. 2019. Streptomycetes: Surrogate hosts for the genetic manipulation of biosynthetic gene clusters and production of natural products. Biotechnol Adv 37:1-20. https://doi.org/10.1016/j.biotechadv.2018.10.003.

2. Qi Y, D'Alessandro JM, Blodgett JA. 2018. Draft Genome Sequence of *Streptomyces* sp. Strain JV178, a Producer of Clifednamide-Type Polycyclic Tetramate Macrolactams. Genome Announc 6: e01401-17. https://doi.org/10.1128/genomeA.01401-17.

3. Blodgett JA, Oh DC, Cao S, Currie CR, Kolter R, Clardy J. 2010. Common biosynthetic origins for polycyclic tetramate macrolactams from phylogenetically diverse bacteria. Proc Natl Acad Sci USA 107:11692–11697. https://doi.org/10.1073/pnas.1001513107.

4. Dhaneesha M, Hasin O, Sivakumar KC, Ravinesh R, Naman CB, Carmeli S, Sajeevan TP. 2019. DNA binding and molecular dynamic studies of polycyclic tetramate macrolactams (PTM) with potential anticancer activity isolated from a sponge-associated *Streptomyces zhaozhouensis* subsp. mycale subsp. nov. Mar Biotechnol 21: 124–137. https://doi.org/10.1007/s10126-018-9866-9.

5. Blodgett JA, Zhang JK, Metcalf WW. 2005. Molecular Cloning, Sequence Analysis, and Heterologous Expression of the Phosphinothricin Tripeptide Biosynthetic Gene Cluster from *Streptomyces viridochromogenes* DSM 40736. Antimicrob Agents Chemother 49:230–240. https://doi.org/10.1128/AAC.49.1.230–240.

6. Chin C, Peluso Paul, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, Cramer GR, Delledonne M, Luo C, Ecker JR, Cantu D, Rank DR, Schatz MC. 2016. Phased diploid genome assembly with single-molecule real-time sequencing. Nat Methods 13:1050–1054. https://doi.org/10.1038/nmeth.4035.

7. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLoS ONE 9: e112963. https://doi.org/10.1371/journal.pone.0112963.

8. Dodt M, Roehr JT, Ahmed R, Dieterich C. 2012. FLEXBAR-Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. Biology 1: 895–905. https://doi.org/10.3390/biology1030895.

9. Li D, Liu C, Luo R, Sadakane K, Lam T. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph. Bioinformatics 31: 1674–1676. https://doi.org/10.1093/bioinformatics/btv033.

10. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch, GD, Shukla M, Thomason JA, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation

pipelines and annotating batches of genomes. Sci Rep 5: 8365. https://doi.org/10.1038/srep08365.

11. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:81–87. https://doi.org/10.1093/nar/gkz310.

12. Rong X, Huang Ying. 2010. Taxonomic evaluation of the *Streptomyces griseus* clade using multilocus sequence analysis and DNA-DNA hybridization, with proposal to combine 29 species and three subspecies as 11 genomic species. Int J Syst Evol Microbiol 60:696-703. https://doi.org/10.1099/ijs.0.012419-0.

13. Edgar, R. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792-1797. https://doi.org /10.1093/nar/gkh340.