AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

| Environmental Microbiology | Announcement

Complete genome sequences of two actinomycetes containing abyssomicin-like gene clusters: *Kutzneria buriramensis* DSM 45791 and *Streptomyces* sp. NL15-2K

Sam E. Williams,[1] David Faurdal,[1] Tue S. Jørgensen,[1] Catherine R. Back,[2] Lynden D. Rooms,[2] Martin A. Hayes,[3] Christine L. Willis,[4] James E. M. Stach,[5] Paul Curnow,[2] Paul R. Race[5]

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT Here, we report the resequencing, assembly, and annotation of two actinomycete genomes containing abyssomicin gene clusters. *Kutzneria buriramensis* DSM 45791 with a circular chromosome of 11,681,598 bp and 4 circular plasmids (14,175–207,548 bp) and *Streptomyces* sp. NL15-2K with a 12,368,159 bp linear genome and circular plasmid (11,584 bp).

KEYWORDS actinomycetes, genomes, abyssomicin, *Streptomyces* NL15-2K, *Kutzneria buriramensis*, genome mining for antibiotics, actinomycetes antibiotic discovery, biosynthetic gene clusters, Oxford Nanopore Sequencing, secondary metabolism

The abyssomicin family of spirotetronate natural products includes potent antibacterial and antitumor agents, whose biosynthesis incorporates uncommon biosynthetic transformations, including Diels-Alderase mediated cycloadditions (1). In search of abyssomicins, we identified two promising but fragmented biosynthetic gene clusters (BGCs) with incomplete assembly: *Streptomyces* NL15-2K, comprising 346 contigs (GCF_003851625.1) and *Kutzneria buriramensis* DSM 45791, comprising 65 contigs (GCF_003387475.1). *Streptomyces* NL15-2K was isolated from forest soil in Vancouver, Canada (2) and studied for lignin degradation (3). *K. buriramensis* DSM 45791 was isolated from forest soil in Buriram, Thailand (4). We resequenced both strains, obtaining *Streptomyces* NL15-2K gDNA (2) from Prof. Motohiro Nishimura (Yasuda Women's University) and *K. buriramensis* from The Leibniz Institute DSMZ. DNA was extracted using Genomic Tip 20/G (QIAGEN) adapted for actinobacteria (5).

Rapid-barcoding libraries (SQK-RBK114-24; Oxford Nanopore Technologies) were sequenced on a MinION with an R10.4 flow cell. Raw sequencing data were demultiplexed and basecalled using Guppy (v6.4.8) high-accuracy model. Illumina, 2 × 250 bp paired-end sequencing was performed on a NovaSeq 6000 (Illumina) using the Nextera XT Library Prep Kit as a service by MicrobesNG (https://microbesng.com/).

All tools were run with default parameters unless otherwise specified. Nanopore reads were trimmed using Filtlong (0.2.1) with min_length 1000 and keep_percent 95 (*K. buriramensis*) and 67.5 (*Streptomyces* NL15-2K). Resulting N50 values were 12,372 bp for *K. buriramensis* and 10,039 bp for *Streptomyces* NL15-2K. Illumina reads were trimmed with Trimmomatic (v0.30) with a Q15 cut-off (6).

Genome assembly for *K. buriramensis* used Trycycler (v0.5.4) (7) with 12 read subsets (read depth 52.4×). Four subsets were assembled by Raven (v1.8.0) (8), Flye (nano-hq) (v2.9.1) (9), or Canu (GenomeSize = 12 m) (v2.1.1) (10). Assemblies were reconciled into five contigs with Trycycler. Streptomyces NL15-2K was assembled using Flye (nano-hq) (v2.9.3), followed by the NPGM-contigger to resolve the terminal inverted repeats (TIR) at

Editor Kenneth M. Stedman, Portland State University, Portland, Oregon, USA

Address correspondence to Sam E. Williams, saedwi@biosustain.dtu.dk.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 28 October 2024
Accepted 19 December 2024
Published 15 January 2025

TABLE 1 Assembly statistics for sequenced genomes and reads after quality filtering steps[d]

Strain	Assembly method	ONT read count (base pairs)	ONT read depth	Contigs per assembly	Final contigs	Illumina paired read count	Illumina read depth[b]	Prepolishing accuracy[c]	Plasmids
Streptomyces NL15-2K	Flye + NPGM-Contigger	64,840 (606,393,990 bp)	48.9×	3[a]	2	2,188,558	75.7×	99.9934%	1
Kutzneria buriramensis DSM 45791	Trycycler	178,781 (1,370,129,713 bp)	113.1×	5–13	5	2,642,996	84.4×	99.9981%	4

[a]Number of contigs before applying the contigger.
[b]Illumina read depth is mean depth across the chromosome contig.
[c]Prepolishing accuracy estimated by polypolish.
[d]ONT, Oxford Nanopore Technologies.

the chromosomal ends (DOI 10.11583/DTU.25133828). Illumina reads were aligned with BWA-MEM2 (v2.2.1) (11) and polished with polypolish (v0.6.0) (12) (Table 1).

The genome of *Streptomyces* NL15-2K has a G + C content of 70.27%, consisting of a 12,368,159 bp linear chromosome with 38 kb TIR and a 11,584 bp circular plasmid. The linear replicon was determined to be complete by mapping all reads against the chromosome and inspecting terminally mapped reads. Final genome coverage was 48.9× ONT and 75.7× Illumina. BUSCO (v5.4.6) (13) analysis (lineage: actinobacteria_class_odb10) showed 100% complete core genes. Phylogenomic analysis (https://tygs.dsmz.de) (14) suggested NL15-2K represents a distinct species with 33.4% DNA-DNA hybridization to its closest relative. NCBI PGAP v.6.7 (15) identified 10,982 genes, 6 complete rRNAs, and 74 tRNAs.

K. buriramensis DSM 45791 has a G + C content of 70.34% across a 11,681,598 bp circular chromosome and 4 plasmids (14,175 bp, 51,178 bp, 56,886 bp, 207,548 bp). Circular chromosome overlaps were resolved with Trycycler, with *dnaA* at the sequence start. Final genome coverage was 113.1× ONT and 84.4× Illumina. BUSCO analysis showed 99.7% complete core genes. NCBI PGAP predicted 11,057 genes, 3 complete rRNAs, 55 tRNAs, and 3 CRISPR arrays.

AntiSMASH 7.1.0 (16) identified 41 BGCs in *Streptomyces* NL15-2K and 51 BGCs in *K. buriramensis*, indicating significant biosynthetic potential.

ACKNOWLEDGMENTS

The authors would like to thank Prof. Motohiro Nishimura from Yasuda Women's University, Hiroshima, Japan, for kindly providing a sample of gDNA for *Streptomyces* NL15-2K.

This work was funded by BBSRC grant BB/T001968/1, a Novo Nordisk Foundation Postdoctoral Fellowship NNF22OC0079021, and through the award of a PhD studentship to L.D.R. (DSTL, DSTLX1000133108).

AUTHOR AFFILIATIONS

[1]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark
[2]School of Biochemistry, University of Bristol, Bristol, United Kingdom
[3]Discovery Sciences, Biopharmaceuticals R&D, AstraZeneca, Sweden, Mölndal
[4]School of Chemistry, University of Bristol, Bristol, United Kingdom
[5]School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom

AUTHOR ORCIDs

Sam E. Williams http://orcid.org/0000-0002-6949-426X
David Faurdal http://orcid.org/0000-0001-8871-1357
Tue S. Jørgensen http://orcid.org/0000-0002-2437-7086

Catherine R. Back http://orcid.org/0000-0002-9835-9171
Martin A. Hayes http://orcid.org/0000-0002-2640-8464
Christine L. Willis http://orcid.org/0000-0002-3919-3642
James E. M. Stach http://orcid.org/0000-0001-8787-2407
Paul Curnow http://orcid.org/0000-0001-9889-0248
Paul R. Race http://orcid.org/0000-0003-0184-5630

FUNDING

Funder	Grant(s)	Author(s)
Novo Nordisk Fonden (NNF)	NNF22OC0079021	Sam E. Williams
UKRI | Biotechnology and Biological Sciences Research Council (BBSRC)	BB/T001968/1	Martin A. Hayes Chris L. Willis Paul R. Race
MOD | Defence Science and Technology Laboratory (Dstl)	DSTLX1000133108	Lynden D. Rooms

AUTHOR CONTRIBUTIONS

Sam E. Williams, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Validation, Visualization, Writing – original draft, Writing – review and editing | David Faurdal, Data curation, Formal analysis, Investigation, Methodology, Writing – review and editing | Tue S. Jørgensen, Data curation, Validation, Writing – review and editing | Catherine R. Back, Resources | Lynden D. Rooms, Resources | Martin A. Hayes, Conceptualization, Writing – review and editing | Christine L. Willis, Conceptualization, Writing – review and editing | James E. M. Stach, Writing – review and editing | Paul Curnow, Project administration, Supervision, Writing – review and editing | Paul R. Race, Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review and editing

DATA AVAILABILITY

All data are available under BioProject PRJNA995553. Raw reads are deposited at SRA with accession SRX21046384–SRX21046385 (Illumina) and SRX21046382–SRX21046383 (ONT). The assembly accession numbers for *K. buriramensis* DSM 45791 are CP144375.1–CP144379.1 and for Streptomyces NL15-2K are CP130630.2 and CP130632.2.

REFERENCES

1. Maschio L, Back CR, Alnawah J, Bowen JI, Johns ST, Mbatha SZ, Han L-C, Lees NR, Zorn K, Stach JEM, Hayes MA, van der Kamp MW, Pudney CR, Burston SG, Willis CL, Race PR. 2024. Delineation of the complete reaction cycle of a natural diels-alderase. Chem Sci 15:11572–11583. https://doi.org/10.1039/d4sc02908a
2. Nishimura M, Kawakami S, Otsuka H. 2019. Draft genome sequence of *Streptomyces* sp. strain NL15-2K, a degrader of lignin-derived aromatic compounds, isolated from forest soil. Microbiol Resour Announc 8:e01456-18. https://doi.org/10.1128/MRA.01456-18
3. Nishimura M, Kawakami S, Otsuka H. 2018. Molecular cloning and characterization of vanillin dehydrogenase from *Streptomyces* sp. NL15-2K. BMC Microbiol 18:154. https://doi.org/10.1186/s12866-018-1309-2
4. Suriyachadkun C, Ngaemthao W, Chunhametha S, Tamura T, Sanglier JJ. 2013. *Kutzneria buriramensis* sp. nov., isolated from soil, and emended description of the genus *Kutzneria*. Int J Syst Evol Microbiol 63:47–52. https://doi.org/10.1099/ijs.0.036533-0
5. Alvarez-Arevalo M, Sterndorff EB, Faurdal D, Jørgensen TS, Mourched AS, Vuksanovic O, Saha S, Weber T. 2023. Extraction and Oxford Nanopore sequencing of genomic DNA from filamentous Actinobacteria. STAR Protoc 4:101955. https://doi.org/10.1016/j.xpro.2022.101955
6. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170
7. Wick RR, Judd LM, Cerdeira LT, Hawkey J, Méric G, Vezina B, Wyres KL, Holt KE. 2021. Trycycler: consensus long-read assemblies for bacterial genomes. Genome Biol 22:266. https://doi.org/10.1186/s13059-021-02483-z
8. Vaser R, Šikić M. 2021. Time- and memory-efficient genome assembly with Raven. Nat Comput Sci 1:332–336. https://doi.org/10.1038/s43588-021-00073-4
9. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8
10. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116
11. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324
12. Wick RR, Holt KE. 2022. Polypolish: Short-read polishing of long-read bacterial genome assemblies. PLoS Comput Biol 18:e1009802. https://doi.org/10.1371/journal.pcbi.1009802

13. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351
14. Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun 10:2182. https://doi.org/10.1038/s41467-019-10210-3
15. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569
16. Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, Fetter A, Terlouw BR, Metcalf WW, Helfrich EJN, van Wezel GP, Medema MH, Weber T. 2023. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res 51:W46–W50. https://doi.org/10.1093/nar/gkad344