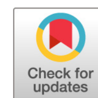
AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology® Resource Announcements

GENOME SEQUENCES



Genome Sequence of *Lysobacter* sp. Strain BMK333-48F3, the Producer Strain of Potent Lipopeptide Antibiotics of the Tripropeptin Family

Patricia Arlt,[a] Hideki Hashizume,[b] Masayuki Igarashi,[b] Harald Gross[a,c]

[a]Department of Pharmaceutical Biology, Pharmaceutical Institute, University of Tübingen, Tübingen, Germany
[b]Department of Microbiology, Institute of Microbial Chemistry, Tokyo, Japan
[c]German Centre for Infection Research, Partner Site Tübingen, Tübingen, Germany

ABSTRACT *Lysobacter* sp. strain BMK333-48F3 is known primarily for its production of the antibiotically active tripropeptins. Here, we report its draft genome sequence, which will give insight into the biosynthesis of tripropeptins and enable genome mining for further secondary metabolites.

Lysobacter sp. strain BMK333-48F3 was isolated from a soil sample collected at Naha, Okinawa Prefecture, Japan (1), and was deposited at the National Institute of Advanced Industrial Science and Technology (accession number FERM BP-7477). It is known for the production of tripropeptin antibiotics (1–6). The members of this lipopeptide class represent potent antibiotics that are highly active against drug-resistant Gram-positive organisms by interfering with cell wall biosynthesis (7, 8). In order to investigate the complete biosynthetic capacity for secondary metabolism, to locate the biosynthetic locus of the tripropeptins, and to clarify the taxonomic position of the strain, the genome of BMK333-48F3 was sequenced at BaseClear (Leiden, The Netherlands).

BMK333-48F3 was harvested from a 1-day culture grown at 27°C and 180 rpm in LB (9), and its genomic DNA (gDNA) was isolated using a gDNA extraction minikit (RBC Bioscience, UK). The gDNA was sheared using a Covaris g-TUBE device; a 10-kb Pacific Biosciences (PacBio) genomic library was generated according to the manufacturer's recommendations, including size selection using the BluePippin size selection system (Sage Science, Inc.), and sequenced on a PacBio RS II instrument using one single-molecule real-time (SMRT) cell. An aliquot of the gDNA obtained was used to create a genomic Nextera XT paired-end library for sequencing using an Illumina HiSeq 2500 platform. The results from both sequencing platforms enabled a *de novo* hybrid assembly. The initial quality assessment of the Illumina data was based on data passing the Illumina chastity filter. Chastity is defined as the brightest base intensity divided by the sum of the brightest and second brightest base intensities. Clusters of reads pass the filter if no more than 1 base call has a chastity value below 0.6 in the first 25 cycles. This filtration process removes the least reliable clusters from the image analysis results (also see reference 10).

Afterward, fastq-mcf, from the ea-utils v1.1.2 package (https://bio.tools/ea-utils), was used for trimming of the Illumina adapters. For phiX removal, Bowtie v2.6.1 (11) was applied to align the reads against the genome of coliphage phiX174 (GenBank accession number NC_001422.1). The unaligned reads were used for downstream analyses. Further quality assessment was based on the remaining reads using FastQC v0.11.5 (12), and low-quality bases were trimmed using BBDuk in BBMap v36.77 (13). The resultant high-quality reads were assembled into contigs using ABySS v2.0.2 (14). The PacBio data were processed and filtered using the SMRT Link software suite, whereby subreads shorter than 50 bp were discarded. Subsequently, the long reads were mapped to the draft assembly using BLASR v1.3.1 (15). These alignments enabled the linkage of contigs, which were in turn placed into

Editor David Rasko, University of Maryland School of Medicine

Copyright © 2021 Arlt et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Harald Gross, harald.gross@uni-tuebingen.de.

The authors declare no conflict of interest.

Received 27 September 2021
Accepted 17 November 2021
Published 9 December 2021

TABLE 1 Sequencing metrics for *Lysobacter* sp. strain BMK333-48F3

Parameter	Finding for *Lysobacter* sp. strain BMK333-48F3
PacBio sequencing	
No. of reads	1,452,682
Mean read length (bp)	3,450
No. of mapped reads	1,061,956
Avg coverage (×)	790
Illumina sequencing	
Read length (nucleotides)	2 × 150
No. of reads	7,259,870
Yield (Mbp)	870
Avg quality score	37.95
Avg coverage (×)	162
Median insert size (bp)	320
De novo hybrid assembly	
Genome size (bp)	5,227,231
GC content (%)	69.7
No. of contigs	7
No. of scaffolds	3
N_{50} (bp)	5,224,492
No. of gaps	4
Avg gap size (bp)	352
Total no. of genes	4,487
No. of coding genes	4,395
No. of predicted biosynthetic gene clusters	12

scaffolds. The orientation and order of and distance between the contigs were estimated using SSPACE-LongRead v1.0 (16), and gapped regions were closed using GapFiller v1.10 (17). Finally, assembly errors and the nucleotide disagreements between the Illumina reads and the scaffold sequences were corrected using Pilon v1.21.1 (18). All software settings were kept at their defaults. Overall, the hybrid genome assembled *de novo* consists of three scaffolds and comprises 5.2 Mb. Annotation was conducted with the Prokaryotic Genome Annotation Pipeline (PGAP) v5.2 pipeline (19, 20), while an automated secondary metabolism analysis was performed using antiSMASH v6.0.1 (21). Essential genome features are summarized in Table 1.

Data availability. The genome sequence has been deposited in the NCBI GenBank database under accession number JAIHOO000000000. The corresponding raw sequencing data sets have been deposited in the NCBI SRA database under accession numbers SRX11855925 (Illumina reads) and SRX11855924 (PacBio reads).

ACKNOWLEDGMENT

We thank the whole-genome sequencing and bioinformatics team from BaseClear BV (Leiden, the Netherlands) for their contributions and ongoing support and for sharing protocols.

REFERENCES

1. Hashizume H, Igarashi M, Hattori S, Hori M, Hamada M, Takeuchi T. 2001. Tripropeptins, novel antimicrobial agents produced by *Lysobacter* sp. I. Taxonomy, isolation and biological activities. J Antibiot (Tokyo) 54:1054–1059. https://doi.org/10.7164/antibiotics.54.1054.
2. Hashizume H, Hirosawa S, Sawa R, Muraoka Y, Ikeda D, Naganawa H, Igarashi M. 2004. Tripropeptins, novel antimicrobial agents produced by *Lysobacter* sp. II. Structure elucidation. J Antibiot (Tokyo) 57:52–58. https://doi.org/10.7164/antibiotics.57.52.
3. Hashizume H, Hirosawa S, Sawa R, Muraoka Y, Ikeda D, Naganawa H, Igarashi M. 2016. Corrigendum: tripropeptins, novel antimicrobial agents produced by *Lysobacter* sp. II. Structure elucidation. J Antibiot (Tokyo) 69:889–891. https://doi.org/10.1038/ja.2016.104.
4. Hashizume H, Hattori S, Igarashi M, Akamatsu Y. 2004. Tripropeptin E, a new tripropeptin group antibiotic produced by *Lysobacter* sp. BMK333-48F3. J Antibiot (Tokyo) 57:394–399. https://doi.org/10.7164/antibiotics.57.394.
5. Hashizume H, Igarashi M, Sawa R, Adachi H, Nishimura Y, Akamatsu Y. 2008. A new type of tripropeptin with anteiso-branched chain fatty acid from *Lysobacter* sp. BMK333-48F3. J Antibiot (Tokyo) 61:577–582. https://doi.org/10.1038/ja.2008.78.
6. Hashizume H, Igarashi M, Sawa R, Adachi H, Nishimura Y, Akamatsu Y. 2016. Corrigendum: a new type of tripropeptin with anteiso-branched chain fatty acid from *Lysobacter* sp. BMK333-48F3. J Antibiot (Tokyo) 69:892–893. https://doi.org/10.1038/ja.2016.105.
7. Hashizume H, Sawa R, Harada S, Igarashi M, Adachi H, Nishimura Y, Nomoto A. 2011. Tripropeptin C blocks the lipid cycle of cell wall biosynthesis by complex formation with undecaprenyl pyrophosphate. Antimicrob Agents Chemother 55:3821–3828. https://doi.org/10.1128/AAC.00443-11.
8. Hashizume H, Takahashi Y, Masuda T, Ohba SI, Ohishi T, Kawada M, Igarashi M. 2018. *In vivo* efficacy of β-lactam/tripropeptin C in a mouse

septicemia model and the mechanism of reverse β-lactam resistance in methicillin-resistant *Staphylococcus aureus* mediated by tripropeptin C. J Antibiot (Tokyo) 71:79–85. https://doi.org/10.1038/ja.2017.88.
9. MacWilliams MP, Liao M-K. 2006. Luria broth (LB) and Luria agar (LA) media and their uses protocol. American Society for Microbiology, Washington, DC.
10. Illumina. 2017. Calculating percent passing filter for patterned and non-patterned flow cells: a comparison of methods for calculating percent passing filter on Illumina flow cells. Illumina, San Diego, CA. https://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/hiseq-x-percent-pf-technical-note-770-2014-043.pdf.
11. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
12. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc.
13. Bushnell B. 2014. BBTools software package. https://sourceforge.net/projects/bbmap.
14. Jackman SD, Vandervalk BP, Mohamadi H, Chu J, Yeo S, Hammond SA, Jahesh G, Khan H, Coombe L, Warren RL, Birol I. 2017. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. Genome Res 27:768–777. https://doi.org/10.1101/gr.214346.116.
15. Chaisson MJ, Tesler G. 2012. Mapping single molecule sequencing reads using Basic Local Alignment with Successive Refinement (BLASR): application and theory. BMC Bioinformatics 13:238. https://doi.org/10.1186/1471-2105-13-238.
16. Boetzer M, Pirovano W. 2014. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. BMC Bioinformatics 15:211. https://doi.org/10.1186/1471-2105-15-211.
17. Boetzer M, Pirovano W. 2012. Toward almost closed genomes with GapFiller. Genome Biol 13:R56. https://doi.org/10.1186/gb-2012-13-6-r56.
18. 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.
19. 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.
20. Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretdin A, Coulouris G, Chitsaz F, Derbyshire MK, Durkin AS, Gonzales NR, Gwadz M, Lanczycki CJ, Song JS, Thanki N, Wang J, Yamashita RA, Yang M, Zheng C, Marchler-Bauer A, Thibaud-Nissen F. 2021. RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. Nucleic Acids Res 49:D1020–D1028. https://doi.org/10.1093/nar/gkaa1105.
21. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res 49:W29–W35. https://doi.org/10.1093/nar/gkab335.