



Metagenomes and Assembled Genomes from Diarrhea-Affected Cattle (*Bos taurus*)

 Tshepisio Pleasure Ateba,^a Kazeem Adekunle Alayande,^b Mulunda Mwanza^a

^aDepartment of Animal Health, Faculty of Natural and Agricultural Sciences, Mafikeng Campus, North West University, Mmabatho, Mafikeng, South Africa

^bDepartment of Microbiology, Faculty of Natural and Agricultural Sciences, Mafikeng Campus, North West University, Mmabatho, Mafikeng, South Africa

ABSTRACT The *de novo* metagenome assembly for C1-TPA is 68,577,389 bp long spread over 10,108 contigs, while that of C3-TPA is 55,517,929 bp distributed over 9,415 contigs. A total of 8 metagenome-assembled genomes (MAGs) were extracted from C1-TPA, and 10 were extracted from C3-TPA. Both samples have a *Flavobacterium* sp. and a *Pseudomonas* sp. in common among their bacterial communities.

Diarrheal disease remains a major cause of morbidity and mortality in the developing world; cattle and young calves are highly susceptible to enteric infections caused by various pathogens (1). Diarrheal samples, C1-TPA and C3-TPA, were collected from affected cattle (*Bos taurus*) directly from the rectum with sterile nitrile gloves, at Lokalleng Village in Mafikeng, South Africa (25.82°S, 25.58°E). About 150 mg of each of the fecal samples was apportioned for metagenome DNA extraction using a Quick-DNA fecal/soil microbe miniprep kit (Zymo Research Corp., USA). The library was prepared with a Nextera DNA Flex library preparation kit (Illumina) using Nextera DNA CD index adapters (96 indexes plated). The final concentrations of the libraries (70.80 ng/ μ l for both C1-TPA and C3-TPA) were measured using the Qubit double-stranded DNA (dsDNA) high-sensitivity (HS) assay kit (Life Technologies), and the average library sizes (521 bp and 523 bp for C1-TPA and C3-TPA, respectively) were determined using the Agilent 2100 bioanalyzer. The libraries were then pooled in equimolar ratios of 8.0 pM and sequenced on an Illumina NovaSeq 6000 system. The numbers of reads generated thereafter were 14,302,284 and 14,431,130 for samples C1-TPA and C3-TPA, respectively. The read length used in the library preparation was 2×150 bp, and the coverage of the sequence was $29\times$ for C1-TPA and $36\times$ for C3-TPA.

The sequenced data were assessed and filtered with Trimmomatic v0.36 (2) for low-quality reads and adapter fragments. The adapter sequences were clipped using a mismatch value of 2, a palindrome clip threshold of 30, and a simple clip threshold of 10. The taxonomy of the metagenomes was determined using Kaiju v1.7.2 (3) and GOTCHA2 v2.1.6. The *de novo* metagenome assembly was constructed with metaSPAdes v3.13.0 (4). Each of the metagenome assemblies was binned using MaxBin 2 v2.2.4 (5) and CONCOCT v1.1 set at different modes—Bowtie2-default and Bowtie2-verysensitive, respectively—and BBMap. The binned contigs were then optimized to exclude bins that have low completeness and high contamination using DAS Tool v1.2. Each bin was then extracted as a metagenome-assembled genome (MAG) and assessed for quality control using CheckM v1.0.18 (6). The taxonomic assignments were obtained for the MAGs based on the genome taxonomy database using GTDB-Tk v1.1.0 (7) and Microbial Genomes Atlas (MiGA) (8); where the taxonomic assignment differs, identity with a higher average nucleotide identity (ANI) percentage was selected.

The MAGs were annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v4.12 (9). The acquired drug-resistant genes were determined using ResFinder v4.0 (10). Most of the software was accessed through the KBase workspace service

Citation Ateba TP, Alayande KA, Mwanza M. 2021. Metagenomes and assembled genomes from diarrhea-affected cattle (*Bos taurus*). *Microbiol Resour Announc* 10:e01411-20. <https://doi.org/10.1128/MRA.01411-20>.

Editor Irene L. G. Newton, Indiana University, Bloomington

Copyright © 2021 Ateba et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Tshepisio Pleasure Ateba, 16100581@nwu.ac.za.

Received 8 December 2020

Accepted 21 January 2021

Published 18 February 2021

TABLE 1 Characteristics and accession numbers of the metagenome-assembled genomes^a

Physical sample	MAG identity	Genome size (bp)	Total CDS	No. of contigs	N_{50} (bp)	G+C content (%)	CMP (%)	CNT (%)	Genome accession no.
C1-TPA	<i>Flavobacterium</i> sp. isolate N1CT	2,649,527	2,357	49	269,215	42.12	99.65	0.00	JADGMX000000000
	<i>Cellulomonas</i> sp. isolate N5CT	4,210,595	3,806	38	144,832	74.00	99.21	1.73	JADGMY000000000
	<i>Brevundimonas</i> sp. isolate N6CT	2,167,354	2,426	375	6,091	70.32	76.06	1.14	JADGMZ000000000
	<i>Salinibacterium</i> sp. isolate N14CT	2,804,916	2,675	23	385,244	68.00	98.55	1.52	JADGNA000000000
	<i>Pseudomonas</i> sp. isolate N17CT	4,778,161	5,202	1151	4,176	62.97	81.61	28.55	JADGNB000000000
	<i>Rhodococcus</i> sp. isolate N19CT	4,327,264	4,079	119	35,912	67.36	97.11	0.88	JADGNC000000000
	<i>Pseudomonas</i> sp. isolate N24CT	4,040,679	3,727	45	149,313	61.02	99.39	3.93	JADGND000000000
	<i>Devosia</i> sp. isolate N26CT	3,978,489	3,820	74	140,542	62.26	98.69	0.00	JADGNE000000000
C3-TPA	<i>Planococcus</i> sp. (in: Bacteria) isolate NTP4	2,417,639	2,663	476	4,504	45.23	62.20	1.88	JADMKI000000000
	<i>Proteiniphilum</i> sp. isolate NTP5	3,297,167	2,664	237	20,538	47.25	98.27	0.70	JADMKJ000000000
	<i>Comamonas</i> sp. isolate NTP6	3,198,324	2,949	61	75,187	57.54	97.33	0.46	JADMKK000000000
	<i>Candidimonas</i> sp. isolate NTP16	3,757,657	3,549	88	79,153	59.04	99.59	0.83	JADMKL000000000
	<i>Pseudomonas stutzeri</i> NTP17	3,293,093	3,121	159	26,172	60.95	95.98	1.23	JADMKM000000000
	<i>Patulibacter</i> sp. isolate NTP18	3,375,244	3,322	297	14,726	71.72	93.16	1.79	JADMKN000000000
	<i>Chryseobacterium</i> sp. isolate NTP27	2,041,705	2,007	187	11,889	39.88	94.00	1.96	JADMKO00000000
	<i>Fermentimonas</i> sp. isolate NTP30	2,897,645	2,309	13	651,827	36.07	100.0	1.64	JADMKP000000000
	<i>Sphingorhabdus</i> sp. isolate NTP38	2,673,275	2,644	243	12,850	54.27	93.14	1.91	JADMKQ00000000
	<i>Flavobacterium</i> sp. isolate NTP45	2,497,823	2,452	275	9,398	43.55	92.90	1.00	JADMKR000000000

^a CMP, completeness; CNT, contamination; CDS, coding sequence. The values for completeness and contamination of each MAG were determined using CheckM v1.0.18, while the genome sizes were determined using v1-KBaseGenomeAnnotations.Assembly-5.0 and the NCBI Prokaryotic Genome Annotation Pipeline (PGAP).

v0.11.1 (11), except for MiGA, PGAP, and ResFinder. Default parameters were used for all software engaged in the analysis except where stated otherwise.

The *de novo* metagenome assembly for C1-TPA has 68,577,389 bp distributed over 10,108 contigs, and the metagenome assembly was binned into eight MAGs. For sample C3-TPA, the assembly size is 55,517,929 bp distributed across 9,415 contigs. Ten different MAGs were extracted from the metagenome. Not all the contigs were binned into the MAGs in both samples. Both samples have *Flavobacterium* spp. (*Flavobacterium* sp. strain N1CT and *Flavobacterium* sp. strain NTP45) and *Pseudomonas* spp. (*Pseudomonas* sp. strain N17CT and *Pseudomonas stutzeri* NTP17) in common (Table 1).

Ethical clearance for the study was approved by the Research Ethics Committee of North West University, South Africa (NWU-00160-14-A9).

Data availability. All data were deposited under the GenBank BioProject number [PRJNA661076](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA661076). The whole-genome shotgun projects have been deposited in DDBJ/ENA/GenBank under the accession numbers [JADGMW000000000](https://www.ncbi.nlm.nih.gov/nuccore/JADGMW000000000) and [JADKLW000000000](https://www.ncbi.nlm.nih.gov/nuccore/JADKLW000000000). The versions described in this paper are the first versions, [JADGMW000000000.1](https://www.ncbi.nlm.nih.gov/nuccore/JADGMW000000000.1) and [JADKLW000000000.1](https://www.ncbi.nlm.nih.gov/nuccore/JADKLW000000000.1). The SRA accession numbers for the raw reads are [SRX9212776](https://www.ncbi.nlm.nih.gov/sra/SRX9212776) and [SRX9218438](https://www.ncbi.nlm.nih.gov/sra/SRX9218438) for samples C1-TPA and C3-TPA, respectively.

ACKNOWLEDGMENTS

This work was supported in part by the Health and Welfare Sector Education and Training Authority (HWSETA) Postgraduate Bursary, South Africa, as well as by an emerging research grant from the Faculty of Natural and Agricultural Sciences, North West University, awarded to T.P.A.

REFERENCES

- El-Ashram S, Aboelhadid SM, Kamel AA, Mahrous LN, Abdelwahab KH. 2019. Diversity of parasitic diarrhea associated with *Buxtonella sulcata* in cattle and buffalo calves with control of buxtonellosis. *Animals* 9:259. <https://doi.org/10.3390/ani9050259>.
- 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>.
- Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun* 7:11257. <https://doi.org/10.1038/ncomms11257>.
- Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 27:824–834. <https://doi.org/10.1101/gr.213959.116>.
- Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32:605–607. <https://doi.org/10.1093/bioinformatics/btv638>.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>.
- Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database.

- Bioinformatics 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>.
8. Rodriguez-R LM, Gunturu S, Harvey WT, Rosselló-Mora R, Tiedje JM, Cole JR, Konstantinidis KT. 2018. The Microbial Genomes Atlas (MiGA) web-server: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Res* 46:W282–W288. <https://doi.org/10.1093/nar/gky467>.
 9. 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>.
 10. Bortolaia V, Kaas RF, Ruppe E, Roberts MC, Schwarz S, Cattoir V, Philippon A, Allesoe RL, Rebelo AR, Florensa AR, Fagelhauer L, Chakraborty T, Neumann B, Werner G, Bender JK, Stingl K, Nguyen M, Coppens J, Xavier BB, Malhotra-Kumar S, Westh H, Pinholt M, Anjum MF, Duggett NA, Kempf I, Nykäsenoja S, Olkkola S, Wiczorek K, Amaro A, Clemente L, Mossong J, Losch S, Ragimbeau C, Lund O, Aarestrup FM. 2020. ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother* 75:3491–3500. <https://doi.org/10.1093/jac/dkaa345>.
 11. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia J-M, Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Devoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, et al. 2018. KBase: the United States Department of Energy Systems Biology Knowledgebase. *Nat Biotechnol* 36:566–569. <https://doi.org/10.1038/nbt.4163>.