Metagenomic Sequencing of Wastewater from a South African Research Farm

Briallen Lobb,[a] Anthony A. Adegoke,[b] Kesen Ma,[a] Andrew C. Doxey,[a] (ID) Olayinka A. Aiyegoro[a,c]

[a]Department of Biology, University of Waterloo, Waterloo, Ontario, Canada
[b]Department of Microbiology, Faculty of Science, University of Uyo, Uyo, Akwa Ibom State, Nigeria
[c]GI Microbiology and Biotechnology Section, Agricultural Research Council–Animal Production, Irene, South Africa

ABSTRACT We sequenced wastewater effluent from the Agricultural Research Council–Animal Production in South Africa that conducts studies on livestock health and farm ecology. *Thauera*, *Oscillibacter*, and *Pseudomonas* were the most abundant genera within the community. Thirty-one different antibiotic resistance genes were identified, 10 of which are associated with tetracycline resistance.

Antibiotics are used to promote growth and manage disease in livestock at Agricultural Research Council–Animal Production in South Africa. However, the spread of antibiotic resistance is a pervasive concern. Waste from farm animals has been shown to spread antibiotic-resistant bacteria, sometimes due to selective pressure found in antibiotic-dosed livestock (1–3). One of a farm's effluents, wastewater, is a documented reservoir of antibiotic resistance genes that could transfer to human pathogens (4–6). Wastewater is also known to contain animal pathogens, some of which are opportunistic and can spread zoonoses (7, 8). Sequencing of the wastewater microbiome can help identify pathogenic species that might exist on the institute's farm and detect antibiotic resistance genes that may be active in these microbial communities.

The metagenome was created from expended water taken from Agricultural Research Council–Animal Production (ARC-AP) in Irene, South Africa. A 1-liter composite sample was created by combining five 200-ml samples collected from different wastewater gutters in the pig facility. The composite sample was centrifuged at 3,500 rpm for 10 min at room temperature to separate the biomass and water. The water was filtered to trap microbes, and DNA was extracted from the pellet on the filter paper. The DNA extraction was done using the FastDNA Spin kit for water (MP Biomedicals, Solon, OH, USA) and the FastPrep apparatus, according to the instructions given by the manufacturer. The DNA was sequenced with the Illumina HiSeq platform and the Illumina HiSeq reagent v3. A total of 28,540,348 read pairs with an average read length of 119 bp each were generated. The reads were trimmed with Sickle version 1.33 (9) and Trim Galore! version 0.5.0 (10) and then assembled using MEGAHIT version 1.1.2 (11), resulting in 58,129 contigs longer than 1 kb. The total assembly length was 311,492,658 bp, with an N_{50} value of 861 bp. Prodigal version 2.6.3 with the -p meta option (12) was used next, facilitating the prediction of 612,922 coding sequences.

A profile of the community based on taxonomic marker genes was constructed with MetAnnotate using the usearch option (13). An abundance for each marker gene hit was calculated using Bowtie 2 version 2.3.4.2 (14), SAMtools version 1.9 (15), and BEDtools version 2.27.1 (16). The abundance of each gene was calculated as the average coverage per base pair across the coding sequence. The most common genera present (based on the average abundance across all taxonomic markers) are *Thauera* (19%), *Oscillibacter* (7%), *Pseudomonas* (6%), *Prevotella* (5%), and *Bacteroides* (3%).

A blastp search of the homolog models in the Comprehensive Antibiotic Resistance

Database (CARD) (17) identified 31 different antibiotic resistance genes that passed CARD's strict score threshold. Ten of these genes are predicted to confer tetracycline resistance, and 5 genes are predicted to confer streptomycin resistance. Annotation with the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (18) using GhostKOALA revealed complete pathways for tetracycline, streptomycin, aminoglycoside, cationic antimicrobial peptide (CAMP), vancomycin, and macrolide resistance, with near-complete pathways for beta-lactam, erythromycin, fluoroquinolone, and lincosamide resistance.

Data availability. This whole-genome shotgun project has been deposited at GenBank under the accession number QXGG00000000. The version described in this paper is version QXGG01000000. The raw reads were deposited in the Sequence Read Archive (SRA) under the accession number SRP159184.

REFERENCES

1. Zhu YG, Johnson TA, Su JQ, Qiao M, Guo GX, Stedtfeld RD, Hashsham SA, Tiedje JM. 2013. Diverse and abundant antibiotic resistance genes in Chinese swine farms. Proc Natl Acad Sci U S A 110:3435–3440. https://doi.org/10.1073/pnas.1222743110.
2. Ghosh S, LaPara TM. 2007. The effects of subtherapeutic antibiotic use in farm animals on the proliferation and persistence of antibiotic resistance among soil bacteria. ISME J 1:191. https://doi.org/10.1038/ismej.2007.31.
3. McNab WW, Singleton MJ, Moran JE, Esser BK. 2007. Assessing the impact of animal waste lagoon seepage on the geochemistry of an underlying shallow aquifer. Environ Sci Technol 41:753–758. https://doi.org/10.1021/es061490j.
4. Zhang Y, Zhang C, Parker DB, Snow DD, Zhou Z, Li X. 2013. Occurrence of antimicrobials and antimicrobial resistance genes in beef cattle storage ponds and swine treatment lagoons. Sci Total Environ 463–464:631–638. https://doi.org/10.1016/j.scitotenv.2013.06.016.
5. Xiong W, Sun Y, Ding X, Wang M, Zeng Z. 2015. Selective pressure of antibiotics on ARGs and bacterial communities in manure-polluted freshwater-sediment microcosms. Front Microbiol 6:194. https://doi.org/10.3389/fmicb.2015.00194.
6. Noyes NR, Yang X, Linke LM, Magnuson RJ, Cook SR, Zaheer R, Yang H, Woerner DR, Geornaras I, McArt JA, Gow SP, Ruiz J, Jones KL, Boucher CA, McAllister TA, Belk KE, Morley PS. 2016. Characterization of the resistome in manure, soil and wastewater from dairy and beef production systems. Sci Rep 6:24645. https://doi.org/10.1038/srep24645.
7. Angulo AB, Nunnery JA, Bair HD, Wint W. 2004. Antimicrobial resistance in zoonotic enteric pathogens. Rev Sci Tech 23:485–496. https://doi.org/10.20506/rst.23.2.1499.
8. Murrell KD, Nawa Y. 1998. Animal waste: risk of zoonotic parasite transmission. Rev Environ Health 13:169–178. https://doi.org/10.1515/REVEH.1998.13.4.169.
9. Joshi NA, Fass JN. 2011. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files. https://github.com/najoshi/sickle.
10. Krueger F. 2015. Trim Galore!: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
11. Li D, Liu CM, Luo R, Sadakane K, Lam TW. 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.
12. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.
13. Petrenko P, Lobb B, Kurtz DA, Neufeld JD, Doxey AC. 2015. MetAnnotate: function-specific taxonomic profiling and comparison of metagenomes. BMC Biol 13:92. https://doi.org/10.1186/s12915-015-0195-4.
14. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357. https://doi.org/10.1038/nmeth.1923.
15. Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27:2987–2993. https://doi.org/10.1093/bioinformatics/btr509.
16. Quinlan AR. 2014. BEDTools: the Swiss-army tool for genome feature analysis. Curr Protoc Bioinformatics 47:11–12. https://doi.org/10.1002/0471250953.bi1112s47.
17. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S, Sharma AN, Doshi S, Courtot M, Lo R, Williams LE, Frye JG, Elsayegh T, Sardar D, Westman EL, Pawlowski AC, Johnson TA, Brinkman FS, Wright GD, McArthur AG. 2017. CARD 2017: expansion and model-centric curation of the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res 45:D566–D573. https://doi.org/10.1093/nar/gkw1004.
18. Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28:27–30. https://doi.org/10.1093/nar/28.1.27.