



Whole-Genome Sequence of *Salmonella enterica* subsp. *enterica* Serovar Typhimurium Strain WG49 and *Escherichia coli* Strain WG5 Used in South Africa for Phage Detection in Water Samples

 L. Bothma,^a D. Gonzalez-Ibeas,^a C. Mienie,^a  C. C. Bezuidenhout,^a  R. Adeleke^b

^aUnit for Environmental Sciences and Management, North-West University, Potchefstroom, South Africa

^bAgricultural Research Council, Institute for Soil, Climate and Water, Pretoria, South Africa

ABSTRACT *Salmonella enterica* subsp. *enterica* serovar Typhimurium WG49 is widely used for enumeration of F-specific RNA (F-RNA) coliphages in water. *Escherichia coli* WG5 is broadly used for the detection and enumeration of somatic coliphages in water samples. We report here the genome sequences of these bacterial strains used in South Africa under ISO methods 10705-1 and 10705-2.

Bacteriophages are biological indicators of fecal contamination of water acting as surrogates for enteric viruses harmful to humans (1–3). Coliphages are phages that infect *Escherichia coli*. Different *E. coli* strains have been used as phage hosts in laboratory assays. Some researchers advocate the use of a specific local strain for each country. However, *E. coli* strain WG5 is broadly used under standardized protocols (ISO method 10705-2 [4]) and is available in the ATCC repository (ATCC 700078). F-Specific RNA (F-RNA) phages are similar to other pathogenic human enteric viruses (e.g., hepatitis A) (3, 5). A method for the enumeration of F-RNA phages was developed by transforming a *Salmonella* strain with a K-12 plasmid coding for the F-pili of *E. coli* (6). Following this approach, *Salmonella* Typhimurium WG49 is broadly used under standardized protocols (ISO method 10705-1 [7]) and is also available in the ATCC repository (ATCC 700730). These strains are used in South Africa for the analysis of water quality in several regions (5, 8, 9) as a reliable and economically affordable method.

Bacterial DNA was isolated with a NucleoSpin tissue kit (Macherey-Nagel) and subjected to Nextera XT DNA library preparation according to Illumina (USA) instructions. Normalized libraries were run on an Illumina MiSeq sequencer (2 × 300-nucleotide [nt] paired-end reads, 1,570,000 reads, 100× estimated coverage). Read quality was evaluated with FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads of quality lower than 35 and lengths less than 45 nt were filtered with Sickle (10). Genome assembly was carried out with SPAdes (11) with default parameters. The assembly was evaluated with QUAST (12).

The *E. coli* WG5 strain has a genome length of 4,513,988 nt, assembled in 151 scaffolds longer than 500 nt and with an N_{50} value of 129,141 nt. Annotation was performed with Rapid Annotation using Subsystems Technology (RAST) (13), giving rise to 4,345 protein coding regions, 31 rRNAs, and 83 tRNAs. Only one region was identified as an incomplete prophage by PHAST (14). Two high-quality and nine questionable clustered regularly interspaced short palindromic repeat (CRISPR) sequences were identified by CRISPRFinder (15). Out of the 20 spacers derived from them, only 1 was found to target the known *Salmonella* phage SJ46. No antibiotic resistance genes were identified by ResFinder (16), but 15 were found by the Antibiotic Resistance Genes Database (ARDB) finder (17).

Received 3 May 2018 Accepted 7 May 2018 Published 28 June 2018

Citation Bothma L, Gonzales-Ibeas D, Mienie C, Bezuidenhout CC, Adeleke R. 2018. Whole-genome sequence of *Salmonella enterica* subsp. *enterica* serovar Typhimurium strain WG49 and *Escherichia coli* strain WG5 used in South Africa for phage detection in water samples. Genome Announc 6:e00372-18. <https://doi.org/10.1128/genomeA.00372-18>.

Copyright © 2018 Bothma et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to C.C. Bezuidenhout, carlos.bezuidenhout@nwu.ac.za.

The *Salmonella* Typhimurium WG49 strain has a 4,868,868-nt length, assembled in 53 scaffolds longer than 500 nt, with an N_{50} value of 299,825 nt. Annotation was performed with RAST (13), giving rise to 4,788 protein coding regions, 31 rRNAs, and 83 tRNAs. Seven regions were identified as potential prophages by PHAST (14), four of them being flagged as “intact,” two as “incomplete,” and one as “questionable.” Three confirmed and two questionable CRISPR sequences were identified by CRISPRFinder (15). Out of the 46 spacers derived from them, similar sequences were found on bacterial genomes falling on annotated CRISPR regions, but none were found to target known sequenced phage genomes. No antibiotic resistance genes were identified by ResFinder (16), but 15 were found by the Antibiotic Resistance Genes Database (ARDB) finder (17).

Accession number(s). The whole-genome shotgun sequencing is described in the GenBank BioProject no. PRJNA434049. Raw reads are available from the Sequence Read Archive (SRA) (accession no. SRR6804879 for *Salmonella* Typhimurium WG49 and SRR6804878 for *E. coli* WG5). The genome sequences have been deposited at DDBJ/ENA/GenBank under accession no. [PXZB00000000](https://doi.org/10.1093/nar/gkm360) for *Salmonella* Typhimurium WG49 and [PYBI00000000](https://doi.org/10.1093/nar/gkm360) for *E. coli* WG5. The versions described in this paper are PXZB01000000 for *Salmonella* Typhimurium WG49 and PYBI01000000 for *E. coli* WG5.

ACKNOWLEDGMENTS

We thank the Water Research Commission (K5/2347//3 and K5/2594//3) and the National Research Foundation (grant no. 109207) for funding. The Agricultural Research Council supported L.B. as part of the Professional Development Program.

The views expressed here are those of the authors and not the funding agencies.

REFERENCES

- AWPRC Study Group. 1991. Bacteriophages as model viruses in water quality control. *Water Res* 25:529–545. [https://doi.org/10.1016/0043-1354\(91\)90126-B](https://doi.org/10.1016/0043-1354(91)90126-B).
- Grabow WOK. 2001. Bacteriophages: update on application as models for viruses in water. *Water SA* 27:251–268. <https://doi.org/10.4314/wsa.v27i2.4999>.
- Leclerc H, Edberg S, Pierzo V, Delattre JM. 2000. Bacteriophages as indicators of enteric viruses and public health risk in groundwaters. *J Appl Microbiol* 88:5–21.
- International Organization for Standardization. 2000. Water quality—detection and enumeration of bacteriophages. Part 2. enumeration of somatic coliphages. International Organization for Standardization, Geneva, Switzerland.
- Nkwe KI, Ateba CN, Sithebe NP, Bezuidenhout CC. 2015. Enumeration of somatic and F-RNA phages as an indicator of fecal contamination in potable water from rural areas of the North West Province. *Pathogens* 4:503–512. <https://doi.org/10.3390/pathogens4030503>.
- Havelaar AH, Hogeboom WM. 1984. A method for the enumeration of male-specific bacteriophages in sewage. *J Appl Bacteriol* 56:439–447. <https://doi.org/10.1111/j.1365-2672.1984.tb01372.x>.
- International Organization for Standardization. 1995. Water quality—detection and enumeration of bacteriophages. Part 1. Enumeration of F-specific RNA bacteriophages. International Organization for Standardization, Geneva, Switzerland.
- South African Bureau of Standards. 2011. Drinking water. Part 1. Microbiological, physical, aesthetic and chemical determinants. South African Bureau of Standards, Pretoria, South Africa.
- Department of Water Affairs and Forestry. 1996. Water quality guidelines for South Africa, 1st ed. Department of Water Affairs and Forestry, Pretoria, South Africa.
- Joshi NA, Fass JN. 2011. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33). Available at <https://github.com/najoshi/sickle>. Retrieved 1 June 2017.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUASt: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 9:75. <https://doi.org/10.1186/1471-2164-9-75>.
- Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. *Nucleic Acids Res* 39:W347–W352. <https://doi.org/10.1093/nar/gkr485>.
- Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a Web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 35:W52–W57. <https://doi.org/10.1093/nar/gkm360>.
- Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 67:2640–2644. <https://doi.org/10.1093/jac/dks261>.
- Liu B, Pop M. 2009. ARDB: antibiotic resistance genes database. *Nucleic Acids Res* 37:D443–D447. <https://doi.org/10.1093/nar/gkn656>.