Supplementary Material

Whole genome characterization of *Listeria welshimeri* isolates recovered from cattle abattoirs and retailers in Gauteng province in South Africa

James Gana[1,2], Rian Ewald Pierneef[3,4,5], Nomakorinte Gcebe[6], Rebone Moerane[1] and Abiodun Adewale Adesiyun[1,7*]

[1]Department of Production Animal Studies, Faculty of Veterinary Science, University of Pretoria, Private Bag X04, Onderstepoort, Pretoria 0110, South Africa; jamesgana38@gmail.com (J.G.); rebone.moerane@up.ac.za (R.M.)

[2]Department of Agricultural Education, Federal College of Education, Kontagora 923101, Niger State, Nigeria

[3]Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria 0001, South Africa; rian.pierneef@up.ac.za

[4]Centre for Bioinformatics and Computational Biology, University of Pretoria, Pretoria 0001, South Africa

[5]Microbiome@UP, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria 0001, South Africa

[6] Bacteriology Department, Onderstepoort Veterinary Research, Agricultural Research Council, Pretoria 0110, South Africa; gceben@arc.agric.za

[7]School of Veterinary Medicine, Faculty of Medical Sciences, University of the West Indies, St. Augustine 685509, Trinidad and Tobago

*****Corresponding author: Abiodun Adewale Adesiyun, e-mail: abiodun.adesiyun@up.ac.za

Materials and methods

Source of *L. welshimeri* isolates

The *L. welshimeri* isolates used in this study were obtained from previous study conducted across 23 cattle farms (328 samples, 0 isolates, 0 %), 8 cattle abattoirs (262 samples, 2 isolates, 0.76 %), and 48 retail outlets (400 samples, 10 isolates, 2.5 %) (Gana et al. 2024a, 2024b).

The types of samples collected, enrichment, use of PCR on enriched broth cultures, plating on selective *Listeria* agar, and identification of *L. monocytogenes*, *L. innocua,* and *L. welshimeri* have been previously described (Gana et al. 2024a; 2024b).

The 12 isolates confirmed as *L. welshimeri* using standard bacteriological methods and multiplex PCR (Gana et al. 2024a; 2024b) were stored at -80 °C until subjected to whole-genome sequencing (WGS).

Variables investigated

The study investigated the effect of selected variables on the frequency and genomic characteristics of *L. welshimeri*, including the following:

the type of farm (communal, cow-calf, and feedlot) and type of feed (grass, grain, and silage) on cattle farms, the size (butcheries, low throughput, and high throughputs) at cattle abattoirs and the size (chain, large, medium, and small) and types of beef and beef products [raw and processed/ready-to-eat (RTE) products]at retail outlets.

DNA extraction from *L. welshimeri* isolates

DNA was extracted from the 12 *L. welshimeri* isolates using the Qiagen DNAEasy Blood & Tissue kit (Thermo Fisher Scientific, Johannesburg, South Africa), according to the manufacturer's instructions.

Whole genome sequencing, genomic analysis, assembly, and annotation

An Illumina MiSeq platform (250-bp paired-end reads; Illumina, Inc., San Diego, CA, USA) was used to sequence all isolates of *L. welshimeri* with the Nextera XT library preparation kit according to the manufacturer's instructions. Quality control, including adapter removal, was conducted using BBDuk (v.38.91; https://jgi.doe.gov/-and-tools/software-tools/bbtools/bb-tools-user-guide/bbdukguide/, accessed on 6 September 2022). SPAdes v.3.15.3 (Bankevich et al. 2012) created a de novo assembly of each isolate with only contigs longer than 500 bp retained for further analysis. The completeness and contamination of the assemblies were assessed with CheckM v.1.1.3 (Parks et al., 2015). The taxonomic classification was achieved using GTDB-Tk v.1.7.0 (Chaumeil et al. 2020). The details are provided in Table S1 and S2 (Supplementary Materials).

In silico MLST

Sequence typing was performed using the MLST tool v2.23.0 (Seemann 2023a), supported by the pubMLST website (https://pubmlst.org/; accessed on 05 August 2024) developed by Jolley and Maiden (2010) and sited at the University of Oxford. The Wellcome Trust funded the development of that website. The latest *Listeria* ST scheme was obtained from BIGSdb-Lm (https://bigsdb.pasteur.fr/listeria/, accessed on 05 August 2024) (Moura et al. 2016) and incorporated into the MLST tool.

Resistance and virulence profiling

The current study used ABRicate (Seemann 2023b) to detect antimicrobial resistance genes and virulence factors. The application was run with default parameters, and the NCBI database was chosen to detect the AMR genes. This database was locally updated and accessed on 18 July 2024 (https://doi.org/10.1128/AAC.00483-19). To detect virulence factors, we used the "vfdb" database, updated and accessed on 18 July 2024 (https://doi.org/10.1093/nar/gkv1239). The virulence profile is based on the presence/absence of a virulence factor in an isolate.

Identification of mobile genetic elements

The presence of plasmids in the genomic data was assessed using geNomad v1.5.1 (Camargo et al. 2024) and the accompanying geNomad database v1.7. The analysis was performed using the "end-to-end" command, and the "-conservative" parameter was evoked to identify plasmids. The results include plasmid-associated conjugation and AMR genes.

Construction of the phylogenomic tree for *L. welshimeri* isolates and correlation with source, type, and class of samples

A phylogenomic tree was constructed using GToTree v1.7.00 (https://github.com/AstrobioMike/GToTree; accessed on 18 July 2024) (Lee 2019) with the HMM profile "Firmicutes" specified (119 target genes). Briefly, Prodigal v2.6.3 (Hyatt et al. 2010) was used to predict genes, and after that, target genes were identified using HMMER3 v3.4 (Eddy 2011). These were individually aligned with Muscle 5.1.linux64 (Edgar 2022), trimmed with TrimAl v1.5.rev0 (Capella-Gutiérrez et al. 2009) and concatenated before phylogenetic estimation with FastTree 2 v2.1.11 (Price et al. 2010). The resultant phylogenomic tree was visualized in R with ggtree (Xu et al. 2020).

Data analysis

The Chi-squared and Fisher's Exact tests for given frequencies were used to determine significant differences, with an alpha level set at 0.05.

References

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19(5):455-477

Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, Chain PSG, Nayfach S, Kyrpides NC (2024) Identification of mobile genetic elements with geNomad. Nature Biotechnol 42(8):1303-1312. https://doi.org/10.1038/s41587-023-01953-y. Epub 2023 Sep 21

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25(15):1972-1973

Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH (2020) GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36(6):1925-1927

Eddy SR (2011) Accelerated profile HMM searches. PLoS Comput Biol 7(10):e1002195

Hyatt D, LoCascio PF, Hauser LJ, Uberbacher EC (2012) Gene and translation initiation site prediction in metagenomic sequences. Bioinformatics 28:2223–2230. https://doi.org/10.1186/1471-2105-11-119

Edgar RC (2022) Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. Nat Commun 13(1):6968

Gana J, Gcebe N, Moerane R, Ngoshe YB, Tshuma, T, Moabelo K, Adesiyun AA (2024a) A comparative study on the occurrence, genetic characteristics, and factors associated with the distribution of *Listeria* species on cattle farms and beef abattoirs in Gauteng Province, South Africa. Trop Anim Health Prod 56:88. https://doi.10.1007/s11250-024-03934-y

Gana J, Gcebe N, Moerane R, Ngoshe YB, Moabelo K, Adesiyun AA (2024b) Detection of pathogenic serogroups and virulence genes in *Listeria monocytogenes* strains isolated from beef and beef products retailed in Gauteng province, South Africa, using phenotypic and polymerase chain reaction (PCR)-based methods. Int J Microbiol 2024:8891963. https://doi.org/10.1155/2024/8891963

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

Jolley KA, Maiden MC (2010) BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11:595

Lee MD (2019) GToTree: a user-friendly workflow for phylogenomics. Bioinformatics 35(20):4162-4164

Moura A, Criscuolo A, Pouseele H, Maury MM, Leclercq A, Tarr C, Björkman JT, Dallman T, Reimer A, Enouf V, Larsonneur E (2016) Whole genome-based population biology and epidemiological surveillance of Listeria monocytogenes. Nat Microbiol 2:16185

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(7):1043-1055

Price MN, Dehal PS, Arkin AP (2010) FastTree 2–approximately maximum-likelihood trees for large alignments. PloS one 5(3):e9490.

Seemann (2023a) - Seemann T, mlst Github https://github.com/tseemann/mlst

Seemann (2023b) - Seemann T, Abricate, Github https://github.com/tseemann/abricate

Xu S, Li L, Luo X, Chen M, Tang W, Zhan L, Dai Z, Lam TT, Guan Y, Yu G (2022) Ggtree: a serialized data object for visualization of a phylogenetic tree and annotation data. IMeta 1(4):e56