Draft Genome Sequence of an *Enterococcus faecalis* Strain Isolated from a Neonatal Blood Sepsis Patient

K. A. Kropp,ᵃ A. Lucid,ᵃ J. Carroll,ᵇ V. Belgrudov,ᶜ P. Walsh,ᵇ B. Kelly,ᵇ C. Smith,ᵈ P. Dickinson,ᵈ A. O'Driscoll,ᶜ K. Templeton,ᵈ,ᵉ P. Ghazal,ᵈ R. D. Sleatorᵃ

Department of Biological Sciences, Cork Institute of Technology, Bishopstown, Cork, Irelandᵃ; NSilico Lifescience Ltd., Cork, Irelandᵇ; Department of Computing, Cork Institute of Technology, Bishopstown, Cork, Irelandᶜ; Division of Pathway Medicine, University of Edinburgh, Edinburgh, Scotland, United Kingdomᵈ; Microbiological Diagnostic Unit, Royal Infirmary, University of Edinburgh, Edinburgh, Scotland, United Kingdomᵉ

Herein, we report the draft genome sequence of *Enterococcus faecalis* ED-NGS-1009, cultivated from a blood sample taken from a neonatal sepsis patient at the Royal Infirmary in Edinburgh, Scotland, United Kingdom.

*E*nterococcus faecalis is a nosocomial, opportunistic pathogen showing resistance to antibiotics, specifically vancomycin (1–4). Preterm neonates are a highly susceptible patient group for bacterial infections (1, 5, 6) and rapid detection of blood sepsis and the causative pathogen are critical steps to enable proper treatment (7–9). Therefore, in the ClouDx-i project, we are aiming to extend our existing knowledge of currently circulating pathogenic strains linked with blood sepsis in neonates to inform the development of new molecular diagnostic assays. Herein, we present the draft genome sequence of an *Enterococcus faecalis* strain isolated from a preterm neonate at the Royal Infirmary in Edinburgh, in 2013. Positivity for blood sepsis and species identification were confirmed by classical microbiological techniques.

The isolate was grown overnight at 37°C on Luria broth (LB) agar, and genomic DNA was isolated using Qiagen genomic tips (Venlo, Limburg, Netherlands). Genomic DNA was fragmented (fragments 2 to 10 kb) using sonication and a non–size selected genome library was produced using the Nextera mate pair kit (Illumina, San Diego, CA). This library was then sequenced on an Illumina MiSeq using MiSeq Reagent kit version 3. Genomic sequence assembly, analysis, and automated reporting were carried out using Simplicity (10). This approach produced 2,954,539 total reads, resulting in an average 222-fold coverage. The average G+C content was 37.27%. Sequence assembly was achieved using a *de novo* assembly pipeline based on the SPAdes version 3.10 assembly tool with k-mers K21, K33, K55, K77, K99, and K127, resulting in a total of 79 contigs, of which 13 were >1,000 bp, representing 99.02% of the total sequence information with the largest contig being 2,285,020 bp in size. Postassembly processing was performed using SPAdes, and only scaffolds of >1,000 bp were considered when estimating genome size as 2,973,146 bp. We annotated the genome with Prokka (11) and used the identified 16S rRNA gene to confirm the species as *Enterococcus faecalis*. A scaffold of the genome was produced with Contiguator2 and we identified the closest related strain by BLASTing the scaffold, returning *E. faecalis* V583 and DENG1 as similar but not identical isolates, as

was evident through insertions and deletions in the genome. The genome was then screened using GLIMMER3 (12) identifying 2,852 open reading frames (ORFs). The predicted ORFs were compared to the UniProt-TrEMBL database (13) using BLASTp, mapping 1,650 ORFs. To identify potential virulence factors in the genome, it was compared to a local database built from the VFDB (14) and Victors databases with BLASTp, using a 75% amino-acid sequence identity cutoff while only considering alignments longer than 100 amino-acids. In all, 63 potential virulence factors were identified.

Samples were handled in accordance with local ethical approval by the ethics committees of the NHS Lothian SAHSC Bioresource and the NHS Research and Development Office, Project ID 2011/R/NE/01 and the HSS BioResource Request ID 13/ES/0126.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number JPWN00000000. The version described in this paper is version JPWN01000000.

REFERENCES

1. **Kaufman D, Fairchild KD.** 2004. Clinical microbiology of bacterial and fungal sepsis in very-low-birth-weight infants. Clin. Microbiol. Rev. **17**: 638–680. http://dx.doi.org/10.1128/CMR.17.3.638-680.2004.
2. **Hasman H, Saputra D, Sicheritz-Ponten T, Lund O, Svendsen CA, Frimodt-Møller N, Aarestrup FM. 30** October 2013. Rapid whole-genome sequencing for the detection and characterization of microorganisms directly from clinical samples. J. Clin. Microbiol. http://dx.doi.org/10.1128/JCM.02452-13.
3. **Fisher K, Phillips C.** 2009. The ecology, epidemiology and virulence of *Enterococcus*. Microbiology **155**:1749–1757. http://dx.doi.org/10.1099/mic.0.026385-0.
4. **Bizzarro MJ, Raskind C, Baltimore RS, Gallagher PG.** 2005. Seventy-five years of neonatal sepsis at Yale: 1928–2003. Pediatrics **116**:595–602. http://dx.doi.org/10.1542/peds.2005-0552.

5. **Shah BA, Padbury JF.** 2014. Neonatal sepsis: an old problem with new insights. Virulence **5:**170–178. http://dx.doi.org/10.4161/viru.26906.

6. **Ghazal P, Dickinson P, Smith CL.** 2013. Early life response to infection. Curr. Opin. Infect. Dis. **26:**213–218. http://dx.doi.org/10.1097/QCO.0b013e32835fb8bf.

7. **Labib AZ, Mahmoud AB, Eissa N, El Gendy FM, Soliman MA, Aly AA.** 2013. Early diagnosis of neonatal sepsis: a molecular approach and detection of diagnostic markers versus conventional blood culture. Int. J. Microbiol. Res. **4:**77–85.

8. **Mancini N, Carletti S, Ghidoli N, Cichero P, Burioni R, Clementi M.** 2010. The era of molecular and other non-culture-based methods in diagnosis of sepsis. Clin. Microbiol. Rev. **23:**235–251. http://dx.doi.org/10.1128/CMR.00043-09.

9. **Sibley CD, Peirano G, Church DL.** 2012. Molecular methods for pathogen and microbial community detection and characterization: current and potential application in diagnostic microbiology. Infect. Genet. Evol. **12:**505–521. http://dx.doi.org/10.1016/j.meegid.2012.01.011.

10. **Walsh P, Carroll J, Sleator RD.** 2013. Accelerating *in silico* research with workflows: a lesson in simplicity. Comput. Biol. Med. **43:**2028–2035. http://dx.doi.org/10.1016/j.compbiomed.2013.09.011.

11. **Seemann T.** 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics **30:**2068–2069. http://dx.doi.org/10.1093/bioinformatics/btu153.

12. **Delcher AL, Harmon D, Kasif S, White O, Salzberg SL.** 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. **27:**4636–4641. http://dx.doi.org/10.1093/nar/27.23.4636.

13. **Boeckmann B, Bairoch A, Apweiler R, Blatter M-C, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M.** 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. **31:**365–370. http://dx.doi.org/10.1093/nar/gkg095.

14. **Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q.** 2005. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res. **33:**D325–D328. http://dx.doi.org/10.1093/nar/gki008.