Complete Genome Assembly of *Enterococcus faecalis* 29212, a Laboratory Reference Strain

T. D. Minogue,ᵃ H. E. Daligault,ᵇ K. W. Davenport,ᵇ S. M. Broomall,ᶜ D. C. Bruce,ᵃ P. S. Chain,ᵇ S. R. Coyne,ᵃ O. Chertkov,ᵇ T. Freitas,ᵇ* H. S. Gibbons,ᶜ J. Jaissle,ᵇ G. I. Koroleva,ᵈ J. T. Ladner,ᵈ G. F. Palacios,ᵈ C. N. Rosenzweig,ᶜ Y. Xu,ᵇ ⓘ S. L. Johnsonᵇ

Diagnostic Systems Division (DSD), United States Army Medical Research Institute of Infectious Diseases (USAMRIID), Fort Detrick, Maryland, USAᵃ; Los Alamos National Laboratory, Los Alamos, New Mexico, USAᵇ; ECBC, Aberdeen Proving Ground, Maryland, USAᶜ; Center for Genome Sciences (CGS), USAMRIID, Fort Detrick, Maryland, USAᵈ

* Present address: T. Freitas, Med Fusion, Lewisville, Texas, USA.

Enterococcus faecalis **is a nonmotile Gram-positive coccus, found both as a commensal organism in healthy humans and animals and as a causative agent of multiple diseases, in particular endocarditis. We sequenced the genome of *E. faecalis* ATCC 29212, a commonly used reference strain in laboratory studies, to complete "finished" annotated assembly (3 Mb).**

Address correspondence to S. L. Johnson, shannonj@lanl.gov.

*E*nterococcus faecalis* can be both a healthy human gut community member and the causative agent of endocarditis, bacteremia, and urinary tract and other infections. The species is often resistant to multiple antibiotics, displaying both inherent and acquired traits (1). *E. faecalis* ATCC strain 29212 was isolated from a human urine sample collected in Portland, Oregon, United States, and is a commonly used laboratory strain for research (it has been noted in over 500 publications).

High-quality genomic DNA was extracted from a purified isolate using a QIAgen Genome Tip-500 at USAMRIID-DSD. Specifically, a 100-mL bacterial culture was grown to stationary phase and nucleic acid extracted as per manufacturer's recommendations. Sequence data included both Illumina and 454 technologies (2, 3). We constructed and sequenced an Illumina "standard" library of 100-bp reads at 300-fold genome-coverage and a separate long insert (8,111 ± 1,087-bp) paired-end library (Roche 454 Titanium). The two libraries were assembled together in Newbler (Roche) and the consensus sequences computationally shredded into 2-kbp overlapping fake reads (shreds). The raw reads were also assembled in Velvet, and those consensus sequences were computationally shredded into 1.5-kbp overlapping shreds (4). Draft data from all platforms were then assembled together with Allpaths and the consensus sequences computationally shredded into 10-kbp overlapping shreds (5). We then integrated the Newbler consensus shreds, Velvet consensus shreds, Allpaths consensus shreds, and a subset of the long-insert read pairs using parallel Phrap (High Performance Software, LLC). Possible mis-assemblies were corrected and some gap closure accomplished with manual editing in Consed (6–8).

Automatic annotation for the *E. faecalis* ATCC 29212 genome utilized an Ergatis based workflow at Los Alamos National Laboratory (LANL) with minor manual curation. The genome consists of one chromosome (2,939,973 bp, 37.5% G+C content) and two plasmids (41,610 bp, 34.3% G+C content, and 66,648 bp, 32.9% G+C content, respectively). Annotation located 2,933 coding sequences, 12 rRNA sequences, and 61 tRNA sequences. All rRNA/tRNA and 96% of the protein coding sequences are located on the chromosome. The annotated genome sequence is publicly available, and the raw data can be provided upon request. Preliminary review of the annotated genome found multiple virulence factors noted by Arias and Murray (1). Further analysis into the intraspecies genetic differences, including this strain, is under way.

Nucleotide sequence accession numbers. The NCBI accession numbers for this genome are CP008816 (chromosome) and CP008814 to CP008815 (plasmids).

REFERENCES

1. **Arias CA, Murray BE.** 2012. The rise of the *Enterococcus*: beyond vancomycin resistance. Nat. Rev. Microbiol. **10**:266–278. http://dx.doi.org/10.1038/nrmicro2761.
2. **Bennett S.** 2004. Solexa Ltd. Pharmacogenomics **5**:433–438. http://dx.doi.org/10.1517/14622416.5.4.433.
3. **Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J-B, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437**:376–380. http://dx.doi.org/10.1038/nature03959.

4. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. **18:**821–829. http://dx.doi.org/10.1101/gr.074492.107.

5. **Butler J, MacCallum I, Kleber M, Shlyakhter IA, Belmonte MK, Lander ES, Nusbaum C, Jaffe DB.** 2008. ALLPATHS: *de novo* assembly of whole-genome shotgun microreads. Genome Res. **18:**810–820. http://dx.doi.org/10.1101/gr.7337908.

6. **Ewing B, Hillier L, Wendl MC, Green P.** 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res. **8:**175–185. http://dx.doi.org/10.1101/gr.8.3.175.

7. **Ewing B, Green P.** 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. Genome Res. **8:**186–194.

8. **Gordon D, Abajian C, Green P.** 1998. Consed: a graphical tool for sequence finishing. Genome Res. **8:**195–202. http://dx.doi.org/10.1101/gr.8.3.195.