



Draft Genome Sequence of *Streptococcus dysgalactiae* subsp. *equisimilis* Strain C161L1 Isolated in Vellore, India

Anshu Babbar, D. Patric Nitsche-Schmitz, Dietmar H. Pieper, Israel Barrantes

Microbial Interactions and Processes Research Group, Helmholtz Centre for Infection Research, Braunschweig, Germany

ABSTRACT *Streptococcus dysgalactiae* subsp. *equisimilis* belongs to the β -hemolytic group C and G pyogenic group of streptococci. Here, we report the draft genome of the *S. dysgalactiae* subsp. *equisimilis* strain C161L1 from Vellore, a region in southern India with a high incidence rate of *S. dysgalactiae* subsp. *equisimilis* infection. This genome is 2.1 Mb long, with a 39.82% G+C content, and encodes 2,022 genes.

Streptococcus dysgalactiae subsp. *equisimilis* is an emerging human pathogen that causes clinical manifestations similar to infections caused by *S. pyogenes* and in some cases has also been implicated in severe invasive infections such as necrotizing fasciitis and streptococcal toxic shock syndrome (1). Here, we present the draft genome of *S. dysgalactiae* subsp. *equisimilis* strain C161L1, which was isolated from throat samples in 2006 from Vellore, a region in southern India that is known to have a relatively high incidence rate of *S. dysgalactiae* subsp. *equisimilis* infections (2).

The genome was sequenced by Illumina MiSeq paired-end sequencing (250-bp per pair), producing 2,325,040 reads and entailing 583,585,040 bp for each sequencing pair. Adapters and other control fragments were removed from the sequencing data after the initial quality check (3). Later, quality trimming was performed using default values for paired-end sequencing (4–7). Subsequently, the trimmed sequencings were assembled using multiple *k*-mer sets (8). This resulted in 319 contigs, which were then annotated (9). During the annotation, two contigs were filtered out for compliance with the public databases. The final assembly comprises 2,121,729 bp and contains no unambiguous bases; contigs over 0.5 kb possess a G+C content of 39.82% and an N_{50} of 61,959 bp (10). These values for genome size and G+C content are in agreement with those of the five complete *S. dysgalactiae* subsp. *equisimilis* genomes in GenBank, which, on average, are 2.13 Mb in size and have a 39.49 to 39.5% G+C content (11). The annotation was carried out using only proteins encoded in existing reference genomes for this species as targets. Here, we found 2,022 genes, distributed among protein-coding genes ($n = 1,972$), tRNAs ($n = 47$), a single tmRNA molecule, and 2 ribosomal RNAs. These values are also within the expected range (11).

As of now, C161L1 is the only representative sequenced from Vellore, India. Therefore, this draft genome will provide insight into the genetic composition and structure of *S. dysgalactiae* subsp. *equisimilis* strains from Vellore, India, and will be useful for comparing and contrasting strains from various geographic locations, especially Europe, where the incidence rates are lower.

Accession number(s). Sequencing reads, assembled contigs, and annotations were deposited in the EMBL-EBI European Nucleotide Archive (ENA) under the accession numbers [FWEH01000001](#) to [FWEH01000317](#).

ACKNOWLEDGMENTS

This work was supported by a bilateral grant of the Indian Council of Medical Research and the German Federal Ministry of Education and Research (project

Received 20 March 2017 Accepted 27 March 2017 Published 11 May 2017

Citation Babbar A, Nitsche-Schmitz DP, Pieper DH, Barrantes I. 2017. Draft genome sequence of *Streptococcus dysgalactiae* subsp. *equisimilis* strain C161L1 isolated in Vellore, India. *Genome Announc* 5:e00336-17. <https://doi.org/10.1128/genomeA.00336-17>.

Copyright © 2017 Babbar et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

Address correspondence to Anshu Babbar, Anshu.Babbar@helmholtz-hzi.de.

01DQ12026) and by the Joint Indo-German Science Centre for Infectious Diseases, HGF project IK-IN001.

REFERENCES

1. Watanabe S, Takemoto N, Ogura K, Miyoshi-Akiyama T. 2016. Severe invasive streptococcal infection by *Streptococcus pyogenes* and *Streptococcus dysgalactiae* subsp. *equisimilis*. *Microbiol Immunol* 60:1–9. <https://doi.org/10.1111/1348-0421.12334>.
2. Reissmann S, Friedrichs C, Rajkumari R, Itzek A, Fulde M, Rodloff AC, Brahmadathan KN, Chhatwal GS, Nitsche-Schmitz DP. 2010. Contribution of *Streptococcus anginosus* to infections caused by groups C and G streptococci, southern India. *Emerg Infect Dis* 16:656–663. <https://doi.org/10.3201/eid1604.090448>.
3. Andrews S. 2010. FastQC a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
4. Buffalo V. 2011. Scythe. <https://github.com/vsbuffalo/scythe>.
5. Joshi N, Fass J. 2011. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files. <https://github.com/najoshi/sickle>.
6. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
7. Mukherjee S, Huntemann M, Ivanova N, Kyrpides NC, Pati A. 2015. Large-scale contamination of microbial isolate genomes by Illumina PhiX control. *Stand Genomic Sci* 10:18. <https://doi.org/10.1186/1944-3277-10-18>.
8. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski 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>.
9. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
10. 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>.
11. Maruyama F, Watanabe T, Nakagawa I. 2016. *Streptococcus pyogenes* genomics, p. 137–203. In Ferretti JJ, Stevens DL, Fischetti VA (ed), *Streptococcus pyogenes: basic biology to clinical manifestations*. University of Oklahoma Health Sciences Center, Oklahoma City, OK.