genomeA
Journals.ASM.org

Genome Sequence of *Streptococcus agalactiae* Strain 09mas018883, Isolated from a Swedish Cow

S. Zubair,[a] E. P. de Villiers,[b] H. H. Fuxelius,[a] G. Andersson,[a] K.-E. Johansson,[c] R. P. Bishop,[b] E. Bongcam-Rudloff[a]

Department of Animal Breeding and Genetics, SLU Global Bioinformatics Centre, Swedish University of Agricultural Sciences, Uppsala, Sweden[a]; International Livestock Research Institute, Nairobi, Kenya[b]; Department of Biomedical Sciences and Veterinary Public Health, Swedish University of Agricultural Sciences, Uppsala, Sweden[c]

We announce the complete genome sequence of *Streptococcus agalactiae* strain 09mas018883, isolated from the milk of a cow with clinical mastitis. The availability of this genome may allow identification of candidate genes, leading to discovery of antigens that might form the basis for development of a vaccine as an alternative means of mastitis control.

Received 24 May 2013 **Accepted** 7 June 2013 **Published** 11 July 2013

Citation Zubair S, de Villiers EP, Fuxelius HH, Andersson G, Johansson K-E, Bishop RP, Bongcam-Rudloff E. 2013. Genome sequence of *Streptococcus agalactiae* strain 09mas018883, isolated from a Swedish cow. Genome Announc. 1(4):e00456-13. doi:10.1128/genomeA.00456-13.

Copyright © 2013 Zubair et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported license.

Address correspondence to S. Zubair, saima.zubair@slu.se.

Streptococcus agalactiae, also known as group B streptococcus (GBS), is a Gram-positive pathogen causing sepsis, meningitis, and pneumonia in human neonates (1) and subclinical mastitis in dairy cattle (2). It infects heifers that have not yet calved and also older cattle in dairy herds in the absence of effective control programs. It is one of the leading causes of economic losses in the dairy industry (3, 4). We have determined the complete genome sequence of an *S. agalactiae* strain, 09mas018883, isolated from bovine (*Bos taurus*) milk.

The genome of *S. agalactiae* 09mas018883 was sequenced by using an Illumina GAIIx employing a paired-end read library strategy with a mean library insert size of 545 bp. After filtering of low-quality reads, MIRA v3.0.0 (5) was used to assemble a total of 10,079,600 quality reads with an average length of 75 bp. The approach of comparative mapping combined with *de novo* assembly was used, followed by gap closure by PCR and Sanger sequencing, GapFiller (6), and Velvet *de novo* assembly (7), ultimately aligning 10,035,130 reads (99.56%) to the genome with an average coverage of 351×. Whole-genome annotation was performed with BASys (8) and RAST (9), and the genome sequence was further analyzed with Artemis and the Artemis Comparison Tool (10).

S. agalactiae 09mas018883 has a circular chromosome of 2,138,694 bp with a G+C content of 35.55%. BASys predicted 2,081 protein-coding genes (CDSs). The open reading frames (ORFs) with putative assigned functions comprise approximately 70%. A total of 80 tRNA genes were predicted by tRNAscan-SE 1.21 (11). rRNA genes were identified by using RNAmmer (12). A total of 21 rRNA genes were predicted, with 7 copies of 16S, 23S, and 5S rRNA genes each.

Genomic islands are the regions of a genomic sequence that vary among closely related strains. These are known to have an important role in the evolution of bacteria by conjugation, transduction, transformation, or horizontal gene transfer, thus underpinning the adaptive capacity of bacteria (13). Genomic islands were identified using IslandViewer (14). Strain 09mas018883 has 7 predicted genomic islands containing 88 genes, including insertion elements, phage genes, prophage genes, membrane-associated proteins, and gene clusters, such as the *purDEK* and *cyl* operons. The *purDEK* operon encodes enzymes that are important for energy metabolism in the *de novo* purine biosynthetic pathway, thus aiding bacterial growth in milk (15). Pilus-like structures are important virulence factors. Genes encoding candidate virulence factors are located in two genomic islands, which each contain at least three genes encoding proteins that contain the conserved amino acid motif LPXTG (16). The 09mas018883 genome contains two type 1 pathogenicity islands, PI-1 and PI-2A. Strain 09mas018883 possesses capsular polysaccharide genes that are also important potential virulence factors. There is an insertion sequence, 2,772 bp in length, containing three predicted genes between the *cpsG* and *cpsH* genes within the *cps* region. The *cps* antigens are also used for the serotyping of GBS isolates (17). The comparative analysis and annotation of the genome are in progress and will be documented in greater detail in a future publication.

Nucleotide sequence accession number. The complete genome sequence of *S. agalactiae* strain 09mas018883 has been deposited in the European Nucleotide Archive (ENA) under the accession no. HF952104.

ACKNOWLEDGMENTS

The work was done under the PhD scholarship program of the Higher Education Commission (HEC) of Pakistan. Sida and Formas supported the research (SIDA reference, 2007–2510; contribution no. 75000506; http://www.sida.se).

The bovine *S. agalactiae* strain 09mas018883 was isolated from a cow with subclinical mastitis in Sweden and was obtained from Helle Unnerstad at the National Veterinary Institute (SVA), Uppsala, Sweden. The genome was sequenced at Science for Life Laboratory sequencing platform, Uppsala, Sweden. The SNIC UPPMAX/UPPNEX next-generation sequencing (NGS) platform supported the handling of large-scale NGS data, and BILS (http://www.bils.se) supported the SLU Global Bioinformatics platform (http://sgbc.slu.se).

REFERENCES

1. **Tettelin H, Masignani V, Cieslewicz MJ, Eisen JA, Peterson S, Wessels MR, Paulsen IT, Nelson KE, Margarit I, Read TD, Madoff LC, Wolf AM, Beanan MJ, Brinkac LM, Daugherty SC, DeBoy RT, Durkin AS, Kolonay JF, Madupu R, Lewis MR, Radune D, Fedorova NB, Scanlan D, Khouri H, Mulligan S, Carty HA, Cline RT, Van Aken SE, Gill J, Scarselli M, Mora M, Iacobini ET, Brettoni C, Galli G, Mariani M, Vegni F, Maione D, Rinaudo D, Rappuoli R, Telford JL, Kasper L, Grandi G, Fraser CM.** 2002. Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*. Proc. Natl. Acad. Sci. U. S. A. **99:**12391–12396.
2. **Radostits OM, Blood DC, Gay CC.** 1997. Veterinary medicine, 8th ed. W. B. Saunders, London, United Kingdom.
3. **Glaser P, Rusniok C, Buchrieser C, Chevalier F, Frangeul L, Msadek T, Zouine M, Couvé E, Lalioui L, Poyart C, Trieu-Cuot P, Kunst F.** 2002. Genome sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease. Mol. Microbiol. **45:**1499–1513.
4. **Erskine RJ.** 1992. Mastitis control in dairy herds with high prevalence of subclinical mastitis. Compend. Contin. Educ. Pract. Vet. **14:**969–979.
5. **Chevreux B, Wetter T, Suhai S.** 1999. Genome sequence assembly using trace signals and additional sequence information. Computer science and biology: proceedings of the German Conference on Bioinformatics **99:** 45–56.
6. **Boetzer M, Pirovano W.** 2012. Toward almost closed genomes with GapFiller. Genome Biol. **13:**56–64.
7. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. **18:**821–829.
8. **Van Domselaar GH, Stothard P, Shrivastava S, Cruz JA, Guo A, Dong X, Lu P, Szafron D, Greiner R, Wishart DS.** 2005. BASys: a web server for automated bacterial genome annotation. Nucleic Acids Res. **33:**455–459.
9. **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. doi:10.1186/1471-2164-9-75.
10. **Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA.** 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics **28:** 464–469.
11. **Lowe TM, Eddy SR.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25:** 955–964.
12. **Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35:**3100–3108.
13. **Juhas M, van der Meer JR, Gaillard M, Harding RM, Hood DW, Crook DW.** 2009. Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol. Rev. **33:**376–393.
14. **Langille MG, Brinkman FS.** 2009. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. Bioinformatics **25:**664–665.
15. **Nilsson D, Kilstrup M.** 1998. Cloning and expression of the *Lactococcus lactis purDEK* genes, required for growth in milk. Appl. Environ. Microbiol. **64:**4321–4327.
16. **Rosini R, Rinaudo CD, Soriani M, Lauer P, Mora M, Maione D, Taddei A, Santi I, Ghezzo C, Brettoni C, Buccato S, Margarit I, Grandi G, Telford JL.** 2006. Identification of novel genomic islands coding for antigenic pilus-like structures in *Streptococcus agalactiae*. Mol. Microbiol. **61:** 126–141.
17. **Ramaswamy SV, Ferrieri P, Madoff LC, Flores AE, Kumar N, Tettelin H, Paoletti LC.** 2006. Identification of novel *cps* locus polymorphisms in nontypable group B *Streptococcus*. J. Med. Microbiol. **55:**775–778.