

# Genome Sequences of Two Pathogenic *Streptococcus agalactiae* Isolates from the One-Humped Camel *Camelus dromedarius*

Saima Zubair,<sup>a</sup> Etienne P. de Villiers,<sup>b</sup> Mario Younan,<sup>d</sup> Göran Andersson,<sup>a</sup> Herve Tettelin,<sup>c</sup> David R. Riley,<sup>c</sup> Joerg Jores,<sup>b</sup> Erik Bongcam-Rudloff,<sup>a</sup> Richard P. Bishop<sup>b</sup>

Department of Animal Breeding and Genetics, SLU Global Bioinformatics Centre, Swedish University of Agricultural Sciences, Uppsala, Sweden<sup>a</sup>; International Livestock Research Institute, Nairobi, Kenya<sup>b</sup>; Institute for Genome Sciences, University of Maryland, School of Medicine, Baltimore, Maryland, USA<sup>c</sup>; Tierärzte ohne Grenzen/Vétérinaires sans Frontières Germany, Nairobi, Kenya<sup>d</sup>

***Streptococcus agalactiae* causes a range of clinical syndromes in camels (*Camelus dromedarius*). We report the genome sequences of two *S. agalactiae* isolates that induce abscesses in Kenyan camels. These genomes provide novel data on the composition of the *S. agalactiae* “pan genome” and reveal the presence of multiple genomic islands.**

Received 10 June 2013 Accepted 13 June 2013 Published 18 July 2013

**Citation** Zubair S, de Villiers EP, Younan M, Andersson G, Tettelin H, Riley DR, Jores J, Bongcam-Rudloff E, Bishop RP. 2013. Genome sequences of two pathogenic *Streptococcus agalactiae* isolates from the one-humped camel *Camelus dromedarius*. *Genome Announc.* 1(4):e00515-13. doi:10.1128/genomeA.00515-13.

**Copyright** © 2013 Zubair et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](http://creativecommons.org/licenses/by/3.0/).

Address correspondence to Saima Zubair, [saima.zubair@slu.se](mailto:saima.zubair@slu.se).

*Streptococcus agalactiae*, also known as group B *Streptococcus* (GBS), is an emerging human pathogen, mainly in neonates (1). *S. agalactiae* infection occurs frequently in camels (*Camelus dromedarius*) and can result in mastitis (2), abscesses, and respiratory tract infections (3). However, the molecular basis of tissue tropism and multiple clinical syndromes is unknown. We have determined and annotated the genome sequences of two different *S. agalactiae* isolates, ILRI005 and ILRI112, associated with abscesses from Kenyan *C. dromedarius*. These pathogenic isolates are genetically distinct according to multilocus sequence typing (A. Fischer, A. M. Liljander, H. Kaspar, C. Muriuki, H. Fuxelius, E. Bongcam-Rudloff, E. deVilliers, C. A. Huber, J. Frey, C. A. Daubenberger, R. Bishop, M. Younan, and J. Jores, submitted for publication).

Sequencing of the *S. agalactiae* isolate ILRI005 was performed using an Illumina genome analyzer (GA) IIx with paired-end-read libraries with a mean library insert size of 210 bp and an average read length of 100 bp. For mapping and *de novo* assembly of 20,687,942 quality reads, we used MIRA v 3.0.0 (4). For mapping we used the genome sequence of the bovine *S. agalactiae* isolate 09mas018883:HF952104 as a reference template (4a). Contigs generated by the *de novo* assembly were ordered using the reference genome, and the consensus genome sequences were aligned using Mauve (5). Genome finishing employed a combination of comparative assembly plus PCR amplification and Sanger sequencing of gaps and GapFiller (6) and Velvet (7) software. A total of 20,189,204 reads (97.56%) were aligned to the reference genome, with an average coverage of 936×. The ILRI112 isolate was sequenced using Ion Torrent technology with a single end-read library with an average read length of 200 bp. Mira v 3.4.1.1 was used to assemble 3,123,413 quality reads. The combination of mapping and *de novo* assembly approaches generated a complete genome sequence with 96% total read alignment and 224× average coverage. Annotation and analysis of the genomes were per-

formed using Basys (8) and RAST (9), Artemis, the Artemis Comparison Tool (10), and Sybil (11).

ILRI005 comprised 2,109,759 bp and ILRI112 2,029,198 bp, with 35.34% and 35.43% GC content, respectively. Identity between homologous regions of the two camel isolate genomes was 99.885% based on JSpeciesv1.2.1 analysis (12). A total of 2,134 open reading frames (ORFs) were predicted in the genome of ILRI005, compared to 2,048 in ILRI112. ILRI005 contained 1,846 genes shared with other *S. agalactiae* strains and 288 unique genes, whereas the ILRI112 genome contained 1,911 shared genes and 137 additional ORFs. Approximately 70% of the predicted ORFs had a putative assigned function.

Genomic islands were predicted using Island Viewer (13). Isolate ILRI005 contained 6 putative genomic islands incorporating 76 predicted genes, whereas ILRI112 contained 7 genomic islands with 117 genes. In contrast to other GBS strains, ILRI005 had an insertion in the region encoding the capsular polysaccharide (cps) (14) of approximately 4,000 bp carrying 8 predicted ORFs, including the *cpsG* and *cpsH* genes.

These genomes provide additional data on the composition of the *S. agalactiae* “pan genome.” Their availability will enable the identification of genes encoding candidate virulence and tissue tropism determinants and the development of specific markers for camel isolates within the type B *Streptococcus* complex.

**Nucleotide sequence accession numbers.** The ILRI005 and ILRI112 genomes have been deposited in ENA under accession numbers [HF952105](https://www.ebi.ac.uk/ena/submitter/ILRI005) and [HF952106](https://www.ebi.ac.uk/ena/submitter/ILRI112).

## ACKNOWLEDGMENTS

This research was supported by funding from GTZ (contract number 81095238) and Swedish International Development and Cooperation (SIDA reference, 2007–2510; contribution no. 75000506; <http://www.sida.se>) and the Higher Education Commission (HEC) of Pakistan.

Camel isolates ILRI005 and ILRI112 were obtained from an abscess lesion (Isiolo Kenya) and a periarticular lesion (Laikipia, Kenya), respec-

tively. The Science for Life Laboratory sequencing platform at Uppsala performed the genome sequencing.

The management and processing of the huge amount of NGS data were supported by the SLU global bioinformatics platform (<http://sgbc.slu.se>), supported by BILS (<http://www.bils.se>), and the SNIC UPPMAX/UPPNEX NGS platform.

## REFERENCES

- Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, DeBoy RT, Davidsen TM, Mora M, Scarselli M, Margarit I, Ros IM, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, O'Connor KJB, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM. 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc. Natl. Acad. Sci. U. S. A.* 102:13950–13955.
- Younan M. 2004. Milk hygiene and udder health, p 67–76. *In* Farah Z, Fischer A (ed), *Milk and meat from the camel*. ETH, Zurich, Switzerland.
- Younan M, Bornstein S, Glücks IV. 2007. Peri-articular abscesses in camel calves in North Kenya. *J. Camel Pract. Res.* 14:161–164.
- Chevreur B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information, p 45–56. *In* *Computer science and biology: proceedings of the German Conference on Bioinformatics*, vol 99. GCB, Göttingen, Germany.
- 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.
- Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. doi:10.1371/journal.pone.0011147.
- Boetzer M, Pirovano W. 2012. Toward almost closed genomes with GapFiller. *Genome Biol.* 13:56–64.
- Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18:821–829.
- 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.
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formisano 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* 9:75–89.
- 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.
- Riley DR, Angiuoli SV, Crabtree J, Dunning Hotopp JC, Tettelin H. 2012. Using Sybil for interactive comparative genomics of microbes on the web. *Bioinformatics* 28:160–166.
- Richter M, Rosselló-Móra R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U. S. A.* 106:19126–19131.
- Langille MG, Brinkman FS. 2009. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics* 25:664–665.
- 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.