THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Draft Genome Sequence of a Streptococcus agalactiae Strain Isolated from a Preterm Neonate Blood Sepsis Patient at the Royal Infirmary, Edinburgh, Scotland

Draft Genome Sequence of a *Streptococcus agalactiae* Strain Isolated from a Preterm Neonate Blood Sepsis Patient at the Royal Infirmary, Edinburgh, Scotland

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. Sleator[a]

Department of Biological Sciences, Cork Institute of Technology, Bishopstown, Cork, Ireland[a]; Nsilico, Cork, Ireland[b]; Department of Computing, Cork Institute of Technology, Bishopstown, Cork, Ireland[c]; Division of Pathway Medicine, University of Edinburgh, Edinburgh, United Kingdom[d]; Microbiological Diagnostic Unit, Royal Infirmary, University of Edinburgh, Edinburgh, United Kingdom[e]

Herein, we report the draft genome sequence of *Streptococcus agalactiae* ED-NGS-1000, cultivated from a blood sample taken from a preterm neonate blood sepsis patient at the Royal Infirmary, Edinburgh, Scotland, United Kingdom.

Address correspondence to R. D. Sleator, roy.sleator@cit.ie.

Streptococcus agalactiae, a member of the group B streptococci, is a Gram-positive pathogen linked to early-onset sepsis in neonates (1–4). Preterm neonates are a highly susceptible patient group, due to their immature immune status and the invasive procedures they are subjected to in neonatal intensive care unit (ICU) settings (1, 5, 6). Rapid detection of blood sepsis and characterization of the causative agent are critical to enable proper treatment (7–9). In the ClouDx-i project, we aim to extend knowledge of currently circulating pathogens linked with neonatal blood sepsis to inform the development of improved molecular diagnostic assays. Herein, we present the draft genome of a *Streptococcus agalactiae* strain isolated from a preterm neonate. Positivity for blood sepsis and identification to the species level were confirmed by classical microbiological identification and characterization techniques.

Following overnight growth at 37°C on Luria broth (LB) agar, genomic DNA was isolated using Qiagen genomic tips (Venlo, Limburg, Netherlands). Genomic DNA fragments, ranging in size from 2 to 10 kb, were produced by sonication. The resulting fragments were then used to produce a non-size-selected genome library with the Nextera mate pair kit (Illumina, San Diego, CA). Libraries were sequenced on an Illumina MiSeq using a MiSeq reagent kit v. 3. Genomic sequence assembly, analysis, and automated reporting were achieved using Simplicity (10). This approach produced 3,295,403 total reads, resulting in an estimated average 307-fold coverage of the genome. The average G+C content was 35.71%. A *de novo* assembly pipeline based on the Spades 3.10 assembly tool was used with k-mers K21, K33, K55, K77, K99, and K127 nucleotides in length, resulting in a total of 135 contigs, of which 8 were >1,000 bp, representing 97.40% of the total sequence information, with the largest contig being 2,181,383 bp in length. The genome was then initially annotated using Prokka (11) and the identified 16S rRNA genes were used to confirm the species as *Streptococcus agalactiae*. A scaffold of the genome was produced with Contiguator2 and we attempted to identify the closest related strain by BLASTing the scaffold against the NCBI database, returning *S. agalactiae* A909 and 09mas018883 as the closest though not absolutely identical sequences, as evidenced by some insertions and deletions in the genome. The genome was then screened using Glimmer 3 (12), revealing 2,125 open reading frames (ORFs). The predicted ORFs were compared to the Uniprot Trembl database (13) using BLASTp. A total of 1,338 ORFs were mapped to the protein database and putative virulence factors were identified by comparison with the VFDB (14) and Victors databases. We used 75% amino acid sequence identity as a cutoff while considering only alignments longer than 100 amino acids, identifying 38 hits.

Samples were handled in accordance with local ethical approval by the ethics committees of the NHS Lothian SAHSC Bioresource and NHS R&D office (project identification [ID] 2011/R/NE/01) and the HSS BioResource (RequestID 13/ES/0126).

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number JPOV00000000. The version described in this paper is version JPOV01000000.

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.** 2014. Rapid whole genome sequencing for the detection and characterization of microorganisms directly from clinical samples. J. Clin. Microbiol. **52:**3136. http://dx.doi.org/10.1128/JCM.01369-14.

3. **Stoll BJ, Hansen N, Fanaroff AA, Wright LL, Carlo WA, Ehrenkranz RA, Lemons JA, Donovan EF, Stark AR, Tyson JE, Oh W, Bauer CR, Korones SB, Shankaran S, Laptook AR, Stevenson DK, Papile LA, Poole WK.** 2002. Changes in pathogens causing early-onset sepsis in very-low-birth-weight infants. N. Engl. J. Med. **347:**240–247. http://dx.doi.org/10.1056/NEJMoa012657.

4. **Stoll BJ, Hansen NI, Sánchez PJ, Faix RG, Poindexter BB, Van Meurs KP, Bizzarro MJ, Goldberg RN, Frantz ID, Hale EC, Shankaran S, Kennedy K, Carlo WA, Watterberg KL, Bell EF, Walsh MC, Schibler K, Laptook AR, Shane AL, Schrag SJ, Das A, Higgins RD; Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network.** 2011. Early onset neonatal sepsis: the burden of group B streptococcal and *E. coli* disease continues. Pediatrics **127:**817–826. http://dx.doi.org/10.1542/peds.2010-2217.

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:**2068–9. 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 MC, 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.