Draft Genome Sequences of *Streptococcus agalactiae* Strains Isolated from Nile Tilapia (*Oreochromis niloticus*) Farms in Thailand

Pattanapon Kayansamruaj,[a] Nopadon Pirarat,[b] Hidehiro Kondo,[c] Ikuo Hirono,[c] Channarong Rodkhum[a]

Department of Veterinary Microbiology, Faculty of Veterinary Science, Chulalongkorn University, Bangkok, Thailand[a]; Department of Veterinary Pathology, Faculty of Veterinary Science, Chulalongkorn University, Bangkok, Thailand[b]; Laboratory of Genome Science, Tokyo University of Marine Science and Technology, Tokyo, Japan[c]

During 2009–2011, two clinical and one environmental strains of *Streptococcus agalactiae* were isolated from Nile tilapia (*Oreochromis niloticus*) farms in Thailand. Draft genome sequences of two clinical isolates comprise 2,048,343 and 2,105,006 bp, while environmental isolates comprise 2,097,115 bp, having 1,573 to 1,578 coding sequences, respectively.

Address correspondence to Channarong Rodkhum, channarong.r@chula.ac.th.

*S*treptococcus agalactiae is the Gram-positive bacterial pathogen that produces serious disease in several mammalian and aquatic animal species (1–3). Several marine and fresh water fishes, including Nile tilapia (*Oreochromis niloticus*), have been reported to be susceptible to *S. agalactiae* infection, causing tremendous economic loss (1, 4). To date, the complete genome sequences of several *S. agalactiae* strains recovered from different countries have been investigated and can be found in the GenBank database. However, the genomic-scaled information of the *S. agalactiae* obtained from the Southeast Asia region is still unavailable. Therefore, the genome sequences of *S. agalactiae* isolated from Nile tilapia farms in Thailand were determined in this study.

Two strains of *S. agalactiae* (FNA07 and FPrA02) were collected from diseased Nile tilapia in 2009 and 2010, while another strain (ENC06) was recovered from water collected from an earthen tilapia pond in 2011. The genome sequencing process was carried out using the Illumina MiSeq platform with 251 paired-end run cycles (Illumina, USA). Once the sequencing process was accomplished, paired-end reads of poor quality were filtered out using the CLC Genomic Workbench version 6.9 (CLC bio, USA) and *de novo* assembly was conducted later. A total of 80 to 106 contigs with 25× average coverage were obtained after the assembly process. The order and orientation of contigs were determined by aligning to *S. agalactiae* strain GD201008-001 (accession number CP003810) using the Contiguator version 2.0 web-based service (5). Draft genome sequences of FNA07, FPrA02, and ENC06 contain 2,105,006, 2,048,343, and 2,097,115 bp, with G+C contents of 35.4, 35.4, and 35.5%, respectively. RNAmmer (6) analysis revealed the number of copies of 5S ($n = 3$; except for FPrA02, $n = 2$), 16S ($n = 1$), and 23S rRNA ($n = 1$) contained in the *S. agalactiae* genomes. Annotation of the genome assemblies was performed using RAST version 4.0 (7), which predicted 2,090, 2,032, and 2,083 coding sequences (CDS). In total, the coding genes were categorized into 341–343 subsystems. Among the CDSs, 19% were responsible for carbohydrate metabolism, 11% for protein metabolism, 9% for cell wall and capsule synthesis, 9% for amino acids and derivatives synthesis, 7% for RNA metabolism, 6% for DNA metabolism, 6% for nucleosides and nucleotides synthesis, and 6% for cofactors, vitamins, prosthetic groups, and pigments. Assigning the proteomes obtained by the RAST server to the OrthoMCL web service (8), about 1,383–1,393 orthologous groups were predicted and approximately 86% of them (1,203–1,205 groups) share common orthologs with other *S. agalactiae* reference strains (accession numbers AE009948, CP000114, CP003810, CP003919, AL732656). The MLST version 1.7 web tool (9) indicated that all 3 strains included in the current study belong to sequence type 7. Prophage identification using the PHAST search tool (10) showed 2 prophage regions (31 and 16 kbp in size) in the FPrA02 and ENC06 genomes, while 3 prophage regions (31, 16 and 8 kbp) were predicted in the FNA07 genome.

Nucleotide sequence accession numbers. The draft genome sequences of *S. agalactiae* strains FNA07, FPrA02, and ENC06 have been deposited at GenBank under the accession numbers JMBI00000000, JMBJ00000000, and JMBK00000000, respectively. The BioProject identification number is PRJNA244605.

ACKNOWLEDGMENTS

REFERENCES

1. **Kayansamruaj P, Pirarat N, Katagiri T, Hirono I, Rodkhum C.** 2014. Molecular characterization and virulence gene profiling of pathogenic *Streptococcus agalactiae* populations from tilapia (*Oreochromis* sp.) farms in Thailand. J. Vet. Diagn. Invest. **26:**488–495. http://dx.doi.org/10.1177/1040638714534237.
2. **Mian GF, Godoy DT, Leal CA, Yuhara TY, Costa GM, Figueiredo HC.** 2009. Aspects of the natural history and virulence of *S. agalactiae* infection in Nile tilapia. Vet. Microbiol. **136:**180–183. http://dx.doi.org/10.1016/j.vetmic.2008.10.016.

3. **Bohnsack JF, Whiting AA, Martinez G, Jones N, Adderson EE, Detrick S, Blaschke-Bonkowsky AJ, Bisharat N, Gottschalk M.** 2004. Serotype III *Streptococcus agalactiae* from bovine milk and human neonatal infections. Emerg. Infect. Dis. **10:**1412–1419. http://dx.doi.org/10.3201/eid1008.030917.

4. **Suanyuk N, Kong F, Ko D, Gilbert GL, Supamattaya K.** 2008. Occurrence of rare genotypes of *Streptococcus agalactiae* in cultured red tilapia *Oreochromis* sp. and Nile tilapia *O. niloticus* in Thailand—relationship to human isolates? Aquaculture **284:**35–40. http://dx.doi.org/10.1016/j.aquaculture.2008.07.034.

5. **Galardini M, Biondi E, Bazzicalupo M, Mengoni A.** 2011. CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes: source code for biology and Medicine. 6:11. http://dx.doi.org/10.1186/1751-0473-6-11.

6. **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. http://dx.doi.org/10.1093/nar/gkm160.

7. **Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R.** 2014. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res. **42:** D206–D214. http://dx.doi.org/10.1093/nar/gkt1226.

8. **Li L, Stoeckert CJ, Jr, Roos DS.** 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. **13:**2178–2189. http://dx.doi.org/10.1101/gr.1224503.

9. **Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM, Lund O.** 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J. Clin. Microbiol. **50:**1355–1361. http://dx.doi.org/10.1128/JCM.06094-11.

10. **Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS.** 2011. PHAST: a fast phage search tool. Nucleic Acids Res. **39:**W347–W352. http://dx.doi.org/10.1093/nar/gkr485.