

# Draft Genome Sequence of *Haemophilus parasuis* gx033, a Serotype 4 Strain Isolated from the Swine Lower Respiratory Tract

Jun Li,<sup>a,b</sup> Hao Peng,<sup>a,b,c</sup> Li-gan Xu,<sup>a,b</sup> Yu-zhou Xie,<sup>a,b</sup> Xiong-biao Xuan,<sup>a,b</sup> Chun-xia Ma,<sup>a,b</sup> Shuai Hu,<sup>a,b</sup> Ze-xiang Chen,<sup>a,b</sup> Wei Yang,<sup>a,b</sup> Yong-ping Xie,<sup>a,b</sup> Yan Pan,<sup>a,b</sup> Li Tao<sup>a,b</sup>

Department of Bacteriology, Guangxi Veterinary Research Institute, Nanning, Guangxi, China<sup>a</sup>; Guangxi Key Laboratory of Animal Vaccines and Diagnostics, Nanning, Guangxi, China<sup>b</sup>; Guangxi University, Nanning, Guangxi, China<sup>c</sup>

J.L. and H.P. contributed equally to this work.

***Haemophilus parasuis* serotype 4 is a Gram-negative pathogen that is the most prevalent *H. parasuis* serovar in the world, but its genome sequence information has not yet been reported. Thus, we determined the genome of *H. parasuis* strain gx033, a serovar 4 strain isolated from a lung specimen of a diseased piglet in southwestern China. Here, we present the first draft genome sequence of this species.**

Received 26 March 2013 Accepted 8 April 2013 Published 23 May 2013

**Citation** Li J, Peng H, Xu L-G, Xie Y-Z, Xuan X-B, Ma C-X, Hu S, Chen Z-X, Yang W, Xie Y-P, Pan Y, Tao L. 2013. Draft genome sequence of *Haemophilus parasuis* gx033, a serotype 4 strain isolated from the swine lower respiratory tract. *Genome Announc.* 1(3):e00224-13. doi:10.1128/genomeA.00224-13.

**Copyright** © 2013 Li 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 Ze-xiang Chen, [xjszxiang@163.com](mailto:xjszxiang@163.com), or Wei Yang, [gyangwei@163.com](mailto:gyangwei@163.com).

*Haemophilus parasuis*, a Gram-negative NAD-dependent rod-shaped bacterium, is generally considered an important pathogen and the etiological agent of Glässer's disease, which is characterized by fibrinous polyserositis, polyarthritis, and meningitis and causes significant financial loss worldwide (1). To date, 15 serovars have been described, using an immunodiffusion test. Serotypes 1, 2, 4, 5, 10, 12, 13, 14, and 15 are considered to be virulent and cause meningitis and pneumonia in swine (2). In North America, serotypes 4 and 5 are the most frequently isolated (3). By combining the results of both the gel diffusion (GD) and the indirect hemagglutination (IHA) test, serovars 4 (24.2%) and 5 (19.2%) were also determined to be the most prevalent serovars in China (4).

Though *H. parasuis* serotype 4 is a predominant pathogen in many countries, its genome sequence information has not been reported. Thus, we sequenced the genome of *H. parasuis* gx033, a serovar 4 strain isolated from a lung specimen of a diseased piglet in southwestern China. The genome sequence of strain gx033 was sequenced with a strategy involving the high-throughput Solexa paired-end sequencing technology (5), using Illumina HiSeq 2000 (Beijing Genomics Institute at Shenzhen, China) (6). A small (500-bp) library and a large (2,000-bp) library were constructed. Sequencing was performed with the pair-end strategy of 90-bp reads to produce 310 Mb of filtered sequences in the 500-bp library and 160 Mb in the 2,000-bp library. The reads were first assembled into 121 contigs with the SOAPdenovo 1.05 (<http://soap.genomics.org.cn/soapdenovo.html>) (7). The contigs were then joined into 35 scaffolds using paired-end information. The genome sequence was analyzed using Glimmer 3.02 (8) for the protein-coding genes, tRNA-scan-SE (9) for tRNA, RNAMmer (10) for rRNA, RepeatMasker and RepeatProteinMasker for transposons, and Tandem Repeats Finder (<http://tandem.bu.edu/trf/trf.html>) for tandem repeat sequences. The functions of the predicted protein-coding genes were then annotated through

comparisons with the Clusters of Orthologous Groups (COG) (11), KEGG (12), and NCBI-NR (13) databases.

The draft genome sequence analysis of gx033 showed a genome size of 2,155,493 bp, with a mean G+C content of 39.79%. In addition, 17,134-bp DNA transposons, 735-bp long interspersed repeated sequences, 2,098-bp long terminal repeated transposons, and 518 short interspersed repeated sequences were found. As is commonly found for other microbial genome sequences, 9.18% of the coding sequences (CDS) correspond to hypothetical proteins of unknown functions. In addition, 116 genes were predicted to encode proteins conferring cell wall, membrane, and envelope biogenesis and 46 genes were related to intracellular trafficking, secretion, and vesicular transport. Furthermore, eight genes related to cell motility were found, such as the gene *pilF* (GL000432), which encodes fimbrial biogenesis and twitching motility protein, which plays a vital role in the adherence to mucosal epithelia in mediating bacteria (14).

This is the first report of the genome sequence of *H. parasuis* serotype 4, and this information should provide further insight into the physiology and metabolic potential of the pathogenesis of the predominant *H. parasuis* serotype; the availability of this genome sequence will provide a better-defined genetic background for future studies of gene expression and regulation.

**Nucleotide sequence accession numbers.** This genome sequence has been deposited at DDBJ/EMBL/GenBank under the accession no. AOSU00000000. The version described in this paper is the first version, accession no. AOSU01000000.

## ACKNOWLEDGMENTS

This work was supported by the Science and Technology research project of the Guangxi Zhuang Autonomous Region Science and Technology Department (0993009-1).

We thank the Beijing Genomics Institute (BGI) for their work in Solexa sequencing, assembly, and annotation of the genome.

## REFERENCES

- Oliveira S, Pijoan C. 2004. Haemophilus parasuis: new trends in diagnosis, epidemiology and control. *Vet. Microbiol.* 68:71–75.
- Vanier G, Szczołka A, Friedl P, Lacouture S, Jacques M, Gottschalk M. 2006. Haemophilus parasuis invades porcine brain microvascular endothelial cells. *Microbiology* 152:135–142.
- Tadjine M, Mittal KR, Bourdon S, Gottschalk M. 2004. Development of a new serological test for serotyping *Haemophilus parasuis* isolates and determination of their prevalence in North America. *J. Clin. Microbiol.* 42:839–840.
- Cai X, Chen H, Blackall PJ, Yin Z, Wang L, Liu Z, Jin M. 2005. Serological characterization of *Haemophilus parasuis* isolates from China. *Vet. Microbiol.* 111:231–236.
- Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IM, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456: 53–59.
- Ding R, Li Y, Qian C, Wu X. 2011. Draft genome sequence of *Paenibacillus elgii* B69, a strain with broad antimicrobial activity. *J. Bacteriol.* 193:4537.
- Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, Li S, Yang H, Wang J, Wang J. 2010. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 20:265–272.
- Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679.
- 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.
- 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.
- Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28:33–36.
- Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32: D277–D280.
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. 2008. GenBank. *Nucleic Acids Res.* 36:D25–D30.
- Watson AA, Alm RA, Mattick JS. 1996. Identification of a gene, *pilF*, required for type 4 fimbrial biogenesis and twitching motility in *Pseudomonas aeruginosa*. *Gene* 180:49–56.