

Complete Genome Sequence of Highly Virulent *Haemophilus parasuis* Serotype 11 Strain SC1401

Ke Dai, Jin Jin, Yiping Wen, Xintian Wen, Lvqin He, Sanjie Cao, Xiaobo Huang, Rui Wu, Qin Zhao

Research Center of Swine Disease, College of Veterinary Medicine, Sichuan Agricultural University, Chengdu, China

K.D. and J.J. contributed equally to this work.

***Haemophilus parasuis*, a normal Gram-negative bacterium, may cause Glässer's disease and pneumonia in pigs. This study aims to identify the genes related to natural competence of the serotype 11 strain SC1401, which frequently shows competence and high pathogenicity. SC1401 shows many differences from strains without natural competence within the molecular basis. We performed complete genome sequencing together with restriction modification system analysis to lay the foundation for later study.**

Received 19 May 2016 Accepted 23 May 2016 Published 21 July 2016

Citation Dai K, Jin J, Wen Y, Wen X, He L, Cao S, Huang X, Wu R, Zhao Q. 2016. Complete genome sequence of highly virulent *Haemophilus parasuis* serotype 11 strain SC1401. *Genome Announc* 4(4):e00628-16. doi:10.1128/genomeA.00628-16.

Copyright © 2016 Dai et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Yiping Wen, yueliang5189@163.com.

Haemophilus parasuis, a typically NAD-dependent commensal bacterium found in the upper respiratory tract of swine (1), can invade the organism and cause a system syndrome called Glässer's disease characterized by polyarthritis, fibrinous polyserositis, and meningitis, resulting in serious economic losses (2). 15 serotypes have been identified, in which serovars 1, 5, 10, 12, 13, and 14 show severe virulence and may cause death within 4 days (3, 4). To date, 22 genomes of *H. parasuis* have been promulgated inclusive of the strain SC1401, which is the first serotype 11 with its complete genome released.

Five strains with experimentally high frequencies of natural competence among 127 wild strains of *H. parasuis* isolated from China were found. Of these five strains, SC1401 was shown to have the relatively highest virulence strain. Serotype identity of SC1401 was confirmed by the Research Center of Swine Disease, College of Veterinary Medicine, Sichuan Agricultural University according to the PCR assay for molecular serotyping on the basis of variation within the capsule loci (5). Screening of bacteria is vital for studies of natural competence.

Whole-genomic DNA was extracted from SC1401 using the TIANamp bacteria DNA kit (Qiagen). The genome of SC1401 was sequenced on Pacbio (Beijing Novogene Bioinformatics Technology Co., Ltd.) by single molecule real-time (SMRT) technology (6) to obtain 38,882 reads and 180× sequencing-depth data. The filtered reads were assembled to generate one contig without gaps. Fifty-nine tRNA genes were predicted with tRNAscan-SE (7), 19 (5s-7,16s-6,23s-6) rRNA genes were predicted with rRNAmmer (8), and 2 small RNAs (sRNAs) were predicted by BLAST against the Rfam (9) database. Repetitive sequences were predicted using RepeatMasker (10) (<http://www.repeatmasker.org/>) to find 111 interspersed repetitive sequences. Tandem repeats were analyzed using Tandem Repeat Finder (11) (<http://www.pathogenomics.sfu.ca/islandviewer/resources.php>) to find 272 tandem repeat sequences. PHAST (12, 13) was used for prophage prediction (<http://phast.wishartlab.com/>) and CRISP-

RFinder (13) for clustered regularly interspaced short palindromic repeat (CRISPR) identification.

Gene prediction was performed on the SC1401 genome assembly by GeneMarks (14) (<http://topaz.gatech.edu/>). To predict protein-coding genes, a whole-genome BLAST search was performed against 6 databases including KEGG, COG, nr, Swiss-Prot, GO, and TrEMBL. Pathogenicity and drug resistance was analyzed searching 4 databases: PHI (pathogen host interactions), VFDB (virulence factors of pathogenic bacteria), ARDB (antibiotic resistance genes), and CAZy (carbohydrate-active enzymes). Secretory proteins were detected on the genome assembly by SignalP (15). Type I-VII secretion system related proteins were extracted from all the annotation results. Type III secretion system effector proteins were detected by EffectiveT3 (16). Secondary metabolite gene clusters were predicted by antiSMASH (17).

The whole-genomic analysis showed a genome size of 2,277,540 bp, with a mean G+C content of 40.03%. The total gene number is 2,234, accounting for 87.65% of the whole genome. Epigenetic modification prediction shows 4,094 loci of type m4C and 23,735 loci of type m6C.

Nucleotide sequence accession numbers. This genome sequence has been deposited at DDBJ/EMBL/GenBank under the accession number [CP015099](https://www.ncbi.nlm.nih.gov/nuccore/CP015099). The version described in this paper is the first version CP015099.1.

ACKNOWLEDGMENTS

We thank Beijing Novogene Bioinformatics Technology Co., Ltd. for their work in sequencing, assembly, and annotation of the genome.

This document is provided for scientific purposes only, not for commercial purpose.

This work was supported by Agro-scientific Research in the Public Interest, grant 201303034.

FUNDING INFORMATION

This work was supported by Agro-scientific Research in the Public Interest grant 201303034.

REFERENCES

- Li J, Peng H, Xu LG, Xie YZ, Xuan XB, Ma CX, Hu S, Chen ZX, Yang W, Xie YP, 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. <http://dx.doi.org/10.1128/genomeA.00224-13>.
- Zhang L, Li Y, Dai K, Wen X, Wu R, Huang X, Jin J, Xu K, Yan Q, Huang Y, Ma X, Wen Y, Cao S. 2015. Establishment of a successive markerless mutation system in *Haemophilus parasuis* through natural transformation. *PLoS One* 10:e0127393. <http://dx.doi.org/10.1371/journal.pone.0127393>.
- Yu Y, Wu G, Zhai Z, Yao H, Lu C, Zhang W. 2015. Fifteen novel immunoreactive proteins of Chinese virulent *Haemophilus parasuis* serotype 5 verified by an immunoproteomic assay. *Folia Microbiol (Praha)* 60:81–87. <http://dx.doi.org/10.1007/s12223-014-0343-1>.
- Angen O, Svensmark B, Mittal KR. 2004. Serological characterization of Danish *Haemophilus parasuis* isolates. *Vet Microbiol* 103:255–258. <http://dx.doi.org/10.1016/j.vetmic.2004.07.013>.
- Howell KJ, Peters SE, Wang J, Hernandez-Garcia J, Weinert LA, Luan SL, Chaudhuri RR, Angen Ø, Aragon V, Williamson SM, Parkhill J, Langford PR, Rycroft AN, Wren BW, Maskell DJ, Tucker AW, BRaDPIT Consortium. 2015. Development of a multiplex PCR assay for rapid molecular serotyping of *Haemophilus parasuis*. *J Clin Microbiol* 53:3812–3821. <http://dx.doi.org/10.1128/JCM.01991-15>.
- Berlin K, Koren S, Chin CS, Drake JP, Landolin JM, Phillippy AM. 2015. Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. *Nat Biotechnol* 33:623–630. <http://dx.doi.org/10.1038/nbt.3238>.
- 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. <http://dx.doi.org/10.1093/nar/25.5.0955>.
- Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, 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>.
- Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Griffiths-Jones S, Eddy SR, Bateman A. 2009. Rfam: updates to the RNA families database. *Nucleic Acids Res* 37:D136–D140. <http://dx.doi.org/10.1093/nar/gkn766>.
- Saha S, Bridges S, Magbanua ZV, Peterson DG. 2008. Empirical comparison of *ab initio* repeat finding programs. *Nucleic Acids Res* 36:2284–2294. <http://dx.doi.org/10.1093/nar/gkn064>.
- Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27:573–580. <http://dx.doi.org/10.1093/nar/27.2.573>.
- 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>.
- Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 35:W52–W57. <http://dx.doi.org/10.1093/nar/gkm360>.
- Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* 29:2607–2618. <http://dx.doi.org/10.1093/nar/29.12.2607>.
- Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8:785–786. <http://dx.doi.org/10.1038/nmeth.1701>.
- Jehl MA, Arnold R, Rattei T. 2011. Effective—a database of predicted secreted bacterial proteins. *Nucleic Acids Res* 39:D591–D595. <http://dx.doi.org/10.1093/nar/gkq1154>.
- Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R. 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res* 39:W339–W346. <http://dx.doi.org/10.1093/nar/gkr466>.