Complete Genome Sequences of Two *Streptococcus suis* Strains Isolated from Asymptomatic Pigs

Atsuko Minowa-Nozawa,* Takashi Nozawa,* Daisuke Takamatsu,*<sup>b</sup> Akemi Yoshida,* Kazunori Murase,* Taisei Kikuchi,* Kasumi Ishida-Kuroki,* Yoshihiro Nitta,* Tsutomu Sekizaki,* Ichiro Nakagawa*

*Department of Microbiology, Graduate School of Medicine, Kyoto University, Kyoto, Japan
*Division of Bacterial and Parasitic Diseases, National Institute of Animal Health, National Agriculture and Food Research Organization, Tsukuba, Ibaraki, Japan
*United Graduate School of Veterinary Sciences, Gifu University, Gifu, Japan
*Laboratory of Genomics, Frontier Science Research Center, University of Miyazaki, Miyazaki, Japan
*Department of Infectious Diseases, Faculty of Medicine, University of Miyazaki, Miyazaki, Japan
*Research Center for Food Safety, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan

**ABSTRACT** *Streptococcus suis* is an important zoonotic pathogen that causes major economic problems in the pig industry worldwide and serious infections in humans, including meningitis and septicemia. Here, we report the complete genome sequences of two strains isolated from asymptomatic pigs.

*Streptococcus suis* is a major swine pathogen that causes various diseases, such as meningitis, septicemia, and endocarditis, which result in large economic losses in the pig industry (1). This bacterium is also an emerging zoonotic pathogen, because life-threatening and fatal *S. suis* infections have been reported in people engaged in slaughtering pigs and processing raw pork (2, 3). On the other hand, a large proportion of healthy adult pigs have been reported to be asymptomatic carriers of *S. suis* (4). DAT299 and DAT300, strains of *S. suis*, are classified as sequence type 114 (ST114) and ST115 in multilocus sequence typing (MLST), respectively, and considered to be weakly pathogenic to pigs and humans (5, 6). Although comparison analysis between pathogenic and weakly pathogenic strains is useful for revelations about the pathogenicity of an organism, whole-genome information of the weakly pathogenic strains is lacking.

Two *S. suis* strains were isolated using Columbia agar supplemented with sheep blood from tonsils of healthy pigs in a meat inspection center at Okinawa, Japan, and shared by the institutions involved in this study. These strains were incubated for 16 h at 37°C in Todd-Hewitt broth supplemented with 2% yeast extract. The bacterial cells were lysed with lysozyme, and then the genomic DNA was extracted (7). Extracted DNA was short- and long-read sequenced on MiSeq (Illumina) and MinION (Oxford Nanopore Technologies [ONT]) platforms, respectively. The Illumina library was prepared using a Nextera DNA library prep kit, and paired-end reads were generated using MiSeq reagent kit v3 (600 cycles, 2 × 300-bp paired-end reads). A MinION library was prepared from unsheared genomic DNA using a rapid barcoding kit (SQK-RBK004) and sequenced with an R9.4.1 flow cell (FLO-MIN106). Base calling with the high-accuracy model and demultiplexing were performed using Guppy (v2.3.5) (ONT), and the quality control was assessed with MinIONC (v1.4.2) (8). The ONT adapters were trimmed using Porechop (v0.2.4) ([https://github.com/rrwick/Porechop](https://github.com/rrwick/Porechop)). The Illumina data were preprocessed using Trimmomatic (v0.36) to remove adapter and low-quality sequences (9) and used in a hybrid assembly with MinION long reads using the Unicycler pipeline (v0.4.7b) with default parameters (10). Genome error correction, circularization, and rotation were also implemented in the Unicycler pipeline. The complete genome
sequence was then annotated using DFAST (v1.2.6) with default parameters (11). Assembly stats, general genome information, and relevant characteristics are summarized in Table 1.

**Data availability.** The complete genome sequences of the two *S. suis* strains in this study have been deposited in DDBJ/EMBL/GenBank under the accession numbers AP023391 and AP023392. The raw Illumina and MinION read data can be found in the DDBJ Sequence Read Archive with the accession numbers DRR239349 to DRR239350 and DRR239357 to DRR239358, respectively.

**ACKNOWLEDGMENTS**

This work was supported in part by grants-in-aid for scientific research (20K07495, 20K15658, 20K21364, 20H05504, 19H03126, 19H03471, 19K06417, 19K07537, 18J40081, and 18K07109) from the Japan Society for the Promotion of Science, Japan Racing and Livestock Promotion Foundation, the Joint Research Project of the Institute of Medical Science, the University of Tokyo, and the Research Program on Emerging and Re-emerging Infectious Diseases (20fk0108073h0003 and 20fk0108130h0401) and by J-PRIDE (19fm0208030h0003) from the Japan Agency for Medical Research and Development (AMED).

**REFERENCES**


### Table 1: Assembly stats, general genome information, and relevant characteristics of *S. suis* strains

<table>
<thead>
<tr>
<th>Strain and genome information</th>
<th>Data for <em>S. suis</em> strain:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DAT299</td>
<td>DAT300</td>
</tr>
<tr>
<td><strong>Strain information</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Origin</td>
<td>Healthy pig</td>
<td>Healthy pig</td>
</tr>
<tr>
<td>Yr of isolation</td>
<td>2005</td>
<td>2006</td>
</tr>
<tr>
<td>ST&lt;sup&gt;a&lt;/sup&gt;</td>
<td>114</td>
<td>115</td>
</tr>
<tr>
<td><strong>Assembly and genome stats</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of reads</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MinION (N&lt;sub&gt;50&lt;/sub&gt; value [bp])</td>
<td>146,234 (8,789)</td>
<td>363,290 (16,687)</td>
</tr>
<tr>
<td>MiSeq</td>
<td>1,029,972</td>
<td>943,120</td>
</tr>
<tr>
<td>Total no. of bases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MinION</td>
<td>274,365,861</td>
<td>1,832,792,270</td>
</tr>
<tr>
<td>MiSeq</td>
<td>290,139,123</td>
<td>266,206,331</td>
</tr>
<tr>
<td>Coverage (&lt;X&gt;)</td>
<td>265.34</td>
<td>904.82</td>
</tr>
<tr>
<td><strong>Chromosome description&lt;sup&gt;b&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genotype size (bp)</td>
<td>2,127,449</td>
<td>2,319,795</td>
</tr>
<tr>
<td>G+C content (%)</td>
<td>41.3</td>
<td>41.1</td>
</tr>
<tr>
<td>No. of CDSs&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2,025</td>
<td>2,207</td>
</tr>
<tr>
<td>Avg protein length (bp)</td>
<td>305.4</td>
<td>305.9</td>
</tr>
<tr>
<td>Coding ratio (%)</td>
<td>87.2</td>
<td>87.3</td>
</tr>
<tr>
<td>No. of rRNAs (rrn operon)</td>
<td>12 (4)</td>
<td>12 (4)</td>
</tr>
<tr>
<td>No. of tRNAs/tmRNAs&lt;sup&gt;d&lt;/sup&gt;</td>
<td>56</td>
<td>56</td>
</tr>
<tr>
<td>No. of CRISPRs</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

<sup>a</sup> ST, sequence type.
<sup>b</sup> All genomic stats are output from DFAST pipeline.
<sup>c</sup> CDSs, coding sequences.
<sup>d</sup> tmRNAs, transfer-messenger RNAs.


