



# Five Complete Genome Sequences Spanning the Dutch *Streptococcus suis* Serotype 2 and Serotype 9 Populations

 Boas C. L. van der Putten,<sup>a,b</sup> Thomas J. Roodsant,<sup>a,b</sup> Martin A. Haagmans,<sup>c</sup> Constance Schultsz,<sup>a,b</sup> Kees C. H. van der Ark<sup>a,b</sup>

<sup>a</sup>Amsterdam UMC, University of Amsterdam, Department of Global Health, Amsterdam Institute for Global Health and Development, Amsterdam, the Netherlands

<sup>b</sup>Amsterdam UMC, University of Amsterdam, Department of Medical Microbiology, Amsterdam, the Netherlands

<sup>c</sup>Amsterdam UMC, University of Amsterdam, Department of Clinical Genetics, Amsterdam, the Netherlands

Boas C. L. van der Putten and Thomas J. Roodsant contributed equally to this work. Author order was determined alphabetically.

**ABSTRACT** The zoonotic pathogen *Streptococcus suis* can cause septicemia and meningitis in humans. We report five complete genomes of *Streptococcus suis* serotype 2 and serotype 9, covering the complete phylogeny of serotype 9 Dutch porcine isolates and zoonotic isolates. The isolates include the model strain S10 and the Dutch emerging zoonotic lineage.

*Streptococcus suis* is an opportunistic pathogen in pigs which can cause zoonotic infections. Human infections are predominantly caused by *S. suis* serotype 2 (1) and can lead to septicemia and meningitis (2). We recently identified a zoonotic *S. suis* serotype 2 clone belonging to clonal complex 20 (CC20), which emerged from a nonzoonotic serotype 9 CC16 clone (3) in the Netherlands. To facilitate further research on the zoonotic potential of *S. suis*, we sequenced the genomes of *S. suis* serotype 9 CC16 and CC20 strains, isolated from diseased pigs, and three serotype 2 strains, including strain S10 (CC1, pig) and two CC20 strains, one each from human and porcine infections (Table 1). Data were generated using Illumina and Nanopore MinION sequencing technologies.

*S. suis* was grown overnight in Todd-Hewitt broth supplemented with yeast extract (THY), and genomic DNA was isolated using the Qiagen MagAttract high-molecular-weight (HMW) DNA extraction kit. The sequence library was constructed using the native barcoding (catalog number EXP-NBD114) and ligation sequencing (catalog number SQK-LSK109) kits (Oxford Nanopore). DNA was repaired and A tailed using NEBNext formalin-fixed, paraffin-embedded (FFPE) DNA repair mix and the NEBNext Ultra II end repair/dA-tailing module (New England BioLabs). A barcode was ligated to the A-tailed DNA using blunt/TA ligase master mix (New England Biolabs). Sequence adapters were ligated to barcoded samples pooled by equal mass with Quick T4 DNA ligase (New England BioLabs). The library was loaded on the flow cell (FLO-MIN106D [R9]) and sequenced using MinKNOW fast base calling version 3.5.5. Default parameters were used for all tools except where noted otherwise. Illumina data were available from our previous study (Table 1) (3).

Illumina read filtering was performed using fastp version 0.20.0 (4). MinION reads were filtered for quality and length using FilTlong version 0.2.0 (5), using the filtered Illumina reads as reference. FastQC version 0.11.8 was used for quality control (6). Illumina and MinION reads were used in a hybrid assembly using Unicycler version 0.4.8, which also performs assembly trimming, circularizing, and rotating (7). Assembly statistics were collected using Quast version 4.6.3 (8). Coverage was assessed using Minimap2 version 2.17 (9), SAMtools version 1.9 (10), and BEDTools version 2.29.0 (11). The complete genomes were annotated using Prokka version 1.14.0 (12). Multilocus sequence typing (MLST) was performed using mlst version 2.17.6 (13). For workflow

**Citation** van der Putten BCL, Roodsant TJ, Haagmans MA, Schultsz C, van der Ark KCH. 2020. Five complete genome sequences spanning the Dutch *Streptococcus suis* serotype 2 and serotype 9 populations. *Microbiol Resour Announc* 9:e01439-19. <https://doi.org/10.1128/MRA.01439-19>.

**Editor** Julie C. Dunning Hotopp, University of Maryland School of Medicine

**Copyright** © 2020 van der Putten 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 Boas C. L. van der Putten, [boas.vanderputten@amsterdamumc.nl](mailto:boas.vanderputten@amsterdamumc.nl), or Kees C. H. van der Ark, [k.vanderark@amsterdamumc.nl](mailto:k.vanderark@amsterdamumc.nl).

**Received** 6 December 2019

**Accepted** 13 January 2020

**Published** 6 February 2020

**TABLE 1** Isolate details, genome information, and accession numbers

Isolate	Isolation source	Serotype	Clonal complex	Genome length (bp)	GC content (%)	No. of total CDSs <sup>a</sup>	Nanopore read N <sub>50</sub> (bp)	Nanopore read No. of Nanopore reads	Nanopore coverage (x)	Nanopore run accession no.	Illumina run accession no.	Assembly accession no.
861160	Human CSF <sup>b</sup>	2	20	2,148,824	41.10	2,029	13,589	7,557	23	ERR3664732	ERR1055554	GCA_902702745
GD-0001	Diseased pig	2	20	2,125,468	41.24	2,014	25,831	10,490	54	ERR3664733	ERR1055586	GCA_902702785
9401240	Diseased pig	9	20	2,195,215	41.43	2,036	11,192	30,418	60	ERR3664735	ERR1055578	GCA_902702775
GD-0088	Diseased pig	9	16	2,298,012	41.20	2,213	7,657	15,321	27	ERR3664734	ERR1055627	GCA_902702765
S10	Diseased pig	2	1	2,048,275	41.32	1,952	15,208	20,251	72	ERR3664731	ERR1055646	GCA_902702755

<sup>a</sup>CDSs, coding sequences.

<sup>b</sup>CSF, cerebrospinal fluid.

management, Snakemake version 5.7.1 (14) was used. The pipeline is freely available from [https://github.com/boasvdp/MRA\\_Streptococcus\\_suis](https://github.com/boasvdp/MRA_Streptococcus_suis).

Genomes of all five strains consisted of a single chromosome ranging from 2,042,889 to 2,292,626 bp with a GC content of 41.10 to 41.43% and a coverage of 23 to 72×, determined using Nanopore data (Table 1).

Draft assemblies of the five strains were 46 to 74 kbp smaller than the complete genomes. Mapping the draft genomes to the complete genomes revealed no missing regions in the draft genomes. The draft genomes are likely smaller than the complete genomes due to the collapse of repeats, which has been described before (15).

**Data availability.** Nanopore, fastq, and fast5 data, as well as the assembled genome sequences, have been deposited in ENA under the accession numbers listed in Table 1 and study number [PRJEB35407](https://www.ncbi.nlm.nih.gov/study/PRJEB35407).

## ACKNOWLEDGMENTS

This study was funded through EU-Horizon2020 grant 727966 (PIGSs) and an Amsterdam UMC Ph.D. grant. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

## REFERENCES

- Huong VT, Ha N, Huy NT, Horby P, Nghia HDT, Thiem VD, Zhu X, Hoa NT, Hien TT, Zamora J, Schultsz C, Wertheim HFL, Hirayama K. 2014. Epidemiology, clinical manifestations, and outcomes of *Streptococcus suis* infection in humans. *Emerg Infect Dis* 20:1105–1114. <https://doi.org/10.3201/eid2007.131594>.
- Wertheim HFL, Nghia HDT, Taylor W, Schultsz C. 2009. *Streptococcus suis*: an emerging human pathogen. *Clin Infect Dis* 48:617–625. <https://doi.org/10.1086/596763>.
- Willemsse N, Howell KJ, Weinert LA, Heuvelink A, Pannekoek Y, Wagenaar JA, Smith HE, van der Ende A, Schultsz C. 2016. An emerging zoonotic clone in the Netherlands provides clues to virulence and zoonotic potential of *Streptococcus suis*. *Sci Rep* 6:28984. <https://doi.org/10.1038/srep28984>.
- Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
- Wick RR. 2019. Filtlong. <https://github.com/rrwick/Filtlong>.
- Andrew S. 2010. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
- Seemann T. 2019. mlst. <https://github.com/tseemann/mlst>.
- Köster J, Rahmann S. 2012. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* 28:2520–2522. <https://doi.org/10.1093/bioinformatics/bts480>.
- Treangen TJ, Salzberg SL. 2011. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet* 13:36–46. <https://doi.org/10.1038/nrg3117>.