





Whole-Genome Sequence of *Streptococcus pyogenes* Strain 591, Belonging to the Genotype *emm49*

 Nadja Patenge,^a  Christian Rückert,^b Jana Bull,^a Kevin Strey,^a Joern Kalinowski,^b Bernd Kreikemeyer^a

^aInstitute of Medical Microbiology, Virology and Hygiene, University Medicine Rostock, Rostock, Germany

^bMicrobial Genomics and Biotechnology, Center for Biotechnology, Universität Bielefeld, Bielefeld, Germany

ABSTRACT *Streptococcus pyogenes* strain 591 is a clinical isolate belonging to the genotype *emm49*. It has been intensively studied for its pathogenicity traits. In this study, the complete genome of strain 591 was sequenced. It consists of a chromosome of 1,762,765 bp with a G+C content of 38.5%.

Streptococcus pyogenes (group A *Streptococcus* [GAS]) is a Gram-positive, exclusively human pathogen responsible for a variety of diseases. Manifestations range from self-limiting superficial infections of the skin or throat to life-threatening invasive diseases. *S. pyogenes* causes up to 18 million invasive infections per year and 500,000 deaths (1). The number of invasive diseases is rising, and untreated superficial infections notoriously lead to the development of severe invasive infections or autoimmune sequelae (2, 3). *S. pyogenes* strain 591 is a patient skin isolate which was kindly provided by R. Lütticken (Aachen, Germany) (4). The strain belongs to the genotype *emm49*, and the molecular basis of its pathogenicity has been extensively characterized. For instance, integrin-dependent invasion into host cells and the role of FCT-3 pilus region-encoded proteins in streptococcal virulence have been studied in this strain (5, 6). Furthermore, it is one of the model strains that have been employed for the identification and functional characterization of small noncoding RNAs (ncRNAs) in *S. pyogenes* (7–9). To facilitate comparative genomic and transcriptomic analyses, the genome of *S. pyogenes* strain 591 was sequenced.

S. pyogenes strain 591 was grown overnight in Todd-Hewitt medium supplemented with 0.2% yeast (THY), and DNA was extracted using the DNeasy blood and tissue kit (Qiagen GmbH, Hilden, Germany). Libraries were prepared using the Nextera XT kit (Illumina, Inc., San Diego, CA), and initial sequencing was conducted using an Illumina MiSeq sequencer (2 × 300-nucleotide [nt] paired-end run). The read quality was checked using FastQC v0.11.3 (10) (option: -t 16). No error correction or adapter trimming were performed other than screening against the phiX library and removing 3' stretches of Ns. The initial assembly of 1,115,170 reads (coverage, 94.3×) using Newbler v2.8 (11) (options: -large, -siom 16, -m, -consed) resulted in 61 contigs in 34 scaffolds. The N_{50} value of the contigs was 62,599 bp. To obtain a complete genome sequence, an additional library for Oxford Nanopore Technologies (ONT) sequencing was prepared using a rapid barcoding kit (SQK-RBK004; Oxford Nanopore Technologies, Oxford, UK) according to the manufacturer's instructions. Sequencing was performed using a R9.5 flow cell on a MinION sequencing platform according to the manufacturer's instructions. Base calling and demultiplexing were performed using Albacore v2.3.1. The obtained ONT data (13,578 reads; coverage, 20.3×) were assembled using Canu v1.6 (12) with the parameter genomeSize = 2m. The read N_{50} value for the ONT data was 9,505 nt. The reads were filtered for size, excluding reads shorter than 1 kb. The read quality was checked using FastQC. Error correction and trimming were performed as part of the Canu assembly. The resulting 8 contigs were polished with the Illumina data using Pilon v1.22 (13). Minimap2 v2.17 (parameters: -ax sr, -secondary = no), BWA-MEM v2 (14) (parameters: -O1, -E1), and Bowtie2 v2.3.2 (15) (parameters: -X 750, -no-unal)

Citation Patenge N, Rückert C, Bull J, Strey K, Kalinowski J, Kreikemeyer B. 2021. Whole-genome sequence of *Streptococcus pyogenes* strain 591, belonging to the genotype *emm49*. Microbiol Resour Announc 10:e00816-21. <https://doi.org/10.1128/MRA.00816-21>.

Editor Steven R. Gill, University of Rochester School of Medicine and Dentistry

Copyright © 2021 Patenge 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 Nadja Patenge, nadja.patenge@med.uni-rostock.de.

Received 17 August 2021

Accepted 13 October 2021

Published 28 October 2021

were used for mapping. Both assemblies were then merged manually using Consed (16). Circularization was performed manually in Consed using the data from the two assemblies. Rotation of the genome was performed using SnapGene (Insightful Science). The resulting gapless genome sequence consisted of 1,762,765 bp, with a G+C content of 38.5%. The final coverage, based on the assembled data, was 102.7×. Automatic annotation using Prokka v1.14.5 (17), followed by manual curation of the Rfam hits, resulted in the assignment of 1,575 genes, out of 1,640 protein-coding sequences, 67 transfer-messenger RNAs (tmRNAs), 18 rRNAs, and 31 noncoding RNAs (ncRNAs). Furthermore, 10 regulatory regions were detected using this approach. These comprised 5 riboswitches, 3 attenuators, and 2 response elements.

Data availability. The complete genome sequence of *S. pyogenes* strain 591 has been deposited in NCBI GenBank and the raw data into the NCBI Sequence Read Archive (SRA). The genome versions described in this paper are the first versions. The accession numbers are PRJNA13286 (BioProject), NZ_CP077685.1 (GenBank), SRR14857120 (SRA; Illumina data), and SRR14857119 (SRA; ONT data).

ACKNOWLEDGMENT

The work of N.P. and K.S. was supported by German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) grant number 420051101.

REFERENCES

- Carapetis JR, Steer AC, Mulholland EK, Weber M. 2005. The global burden of group A streptococcal diseases. *Lancet Infect Dis* 5:685–694. [https://doi.org/10.1016/S1473-3099\(05\)70267-X](https://doi.org/10.1016/S1473-3099(05)70267-X).
- Cunningham MW. 2008. Pathogenesis of group A streptococcal infections and their sequelae. *Adv Exp Med Biol* 609:29–42. https://doi.org/10.1007/978-0-387-73960-1_3.
- Sims SA, Colquhoun S, Wyber R, Carapetis JR. 2016. Global disease burden of group A *Streptococcus*. doi:NBK333415 [bookaccession].
- Kaufhold A, Podbielski A, Johnson DR, Kaplan EL, Luttkicken R. 1992. M protein gene typing of *Streptococcus pyogenes* by nonradioactively labeled oligonucleotide probes. *J Clin Microbiol* 30:2391–2397. <https://doi.org/10.1128/jcm.30.9.2391-2397.1992>.
- Nakata M, Koller T, Moritz K, Ribardo D, Jonas L, McIver KS, Sumitomo T, Terao Y, Kawabata S, Podbielski A, Kreikemeyer B. 2009. Mode of expression and functional characterization of FCT-3 pilus region-encoded proteins in *Streptococcus pyogenes* serotype M49. *Infect Immun* 77:32–44. <https://doi.org/10.1128/IAI.00772-08>.
- Siemens N, Patenge N, Otto J, Fiedler T, Kreikemeyer B. 2011. *Streptococcus pyogenes* M49 plasminogen/plasmin binding facilitates keratinocyte invasion via integrin-integrin-linked kinase (ILK) pathways and protects from macrophage killing. *J Biol Chem* 286:21612–21622. <https://doi.org/10.1074/jbc.M110.202671>.
- Khani A, Popp N, Kreikemeyer B, Patenge N. 2018. A glycine riboswitch in *Streptococcus pyogenes* controls expression of a sodium:alanine symporter family protein gene. *Front Microbiol* 9:200. <https://doi.org/10.3389/fmicb.2018.00200>.
- Pappesch R, Warnke P, Mikkat S, Normann J, Wisniewska-Kucper A, Huschka F, Wittmann M, Khani A, Schwengers O, Oehmcke-Hecht S, Hain T, Kreikemeyer B, Patenge N. 2017. The regulatory small RNA MarS supports virulence of *Streptococcus pyogenes*. *Sci Rep* 7:12241. <https://doi.org/10.1038/s41598-017-12507-z>.
- Patenge N, Billion A, Raasch P, Normann J, Wisniewska-Kucper A, Retej J, Boisguerin V, Hartsch T, Hain T, Kreikemeyer B. 2012. Identification of novel growth phase- and media-dependent small non-coding RNAs in *Streptococcus pyogenes* M49 using intergenic tiling arrays. *BMC Genomics* 13:550. <https://doi.org/10.1186/1471-2164-13-550>.
- Wingett SW, Andrews S. 2018. FastQ Screen: a tool for multi-genome mapping and quality control. *F1000Res* 7:1338. <https://doi.org/10.12688/f1000research.15931.2>.
- Miller JR, Koren S, Sutton G. 2010. Assembly algorithms for next-generation sequencing data. *Genomics* 95:315–327. <https://doi.org/10.1016/j.ygeno.2010.03.001>.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* 27:722–736. <https://doi.org/10.1101/gr.215087.116>.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 1303.3997v2 [q-bio.GN]. <https://arxiv.org/abs/1303.3997>.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
- Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Res* 8:195–202. <https://doi.org/10.1101/gr.8.3.195>.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.