Complete Genome Sequence and Annotation for *Turicibacter sanguinis* MOL361<sup>T</sup> (DSM 14220)

© Joel J. Maki, a,b,c Daniel W. Nielsen, a,c Torey Looft a

a Food Safety and Enteric Pathogens Research Unit, National Animal Disease Center, Agricultural Research Service, U.S. Department of Agriculture, Ames, Iowa, USA  
b Interdepartmental Microbiology Graduate Program, Iowa State University, Ames, Iowa, USA  
c Oak Ridge Institute for Science and Education, ARS Research Participation Program, Oak Ridge, Tennessee, USA  

**ABSTRACT**  *Turicibacter sanguinis* MOL361 (DSM 14220) is the reference and type strain for the *Turicibacter* genus, commonly found in the intestinal tract of animal species. Long-read sequencing was performed on this strain to complement publicly available Illumina HiSeq-based data, producing a complete annotated genome sequence.

*Turicibacter* is a genus of the *Firmicutes* phylum, commonly identified in the intestine of animals (1–8). *Turicibacter* species are capable of interaction with host-derived compounds, suggesting that they contribute to physiological processes (9–13). Here, we describe the complete genome sequence of *Turicibacter sanguinis* MOL361<sup>T</sup>.

*Turicibacter sanguinis* MOL361 is the type strain for the type species of *Turicibacter* and was sequenced in 2019 as part of the Department of Energy (DOE) Joint Genome Institute project Genomic Encyclopedia of Bacterial and Archaeal Type Strains, Phase IV: the One Thousand Microbial Genomes (KMG IV). The incomplete draft genome was uploaded to NCBI (assembly accession number GCA_004338625.1). To complete the *T. sanguinis* genome, long-read sequence data were generated for assembly with existing short-read data. *T. sanguinis* DSM 14220 was obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ) and grown anaerobically at 37°C on brain heart infusion (BHI) broth prior to sequencing. Genomic DNA was isolated using the PureLink genomic DNA extraction minikit (Invitrogen, Waltham, MA) according to the manufacturer’s instructions. The extracted DNA was assessed using gel electrophoresis, a Qubit fluorimeter (double-stranded DNA [dsDNA] broad-range [BR] kit; Life Technologies, Waltham, MA), and a NanoDrop instrument (Thermo Fisher Scientific, Waltham, MA) prior to sequencing. The genomic library for Nanopore sequencing was prepared with the SQK-RBK004 rapid barcoding kit (Oxford Nanopore, Oxford, UK) according to the manufacturer’s instructions.

Default parameters were used for all software unless otherwise noted. A FLOMIN106 (R9.4.1) flow cell was used on a MinION device (Oxford Nanopore) for sequencing of the *T. sanguinis* MOL361<sup>T</sup> genomic DNA. The flow cell was run for 48 h, and reads were quality (Q) scored (Q ≥ 7), demultiplexed, and trimmed with Guppy v3.1.5 (14). A total of 66,759 MinION reads (genomic coverage, 30×) with a mean length of 2,632 bp (N<sub>50</sub>, 4,949 bp) were generated. Paired-end Illumina HiSeq reads (2 × 150 bp; genomic coverage, 240×) were downloaded from GenBank’s Sequence Read Archive (SRA) (accession number PRJNA500327) using the SRA Toolkit (15). Illumina and Nanopore reads were assembled into 2 contigs with Unicycler v0.4.7 in “bold” mode (16). Geneious Prime v2019.1.1 (Biomatters, Ltd., Auckland, New Zealand) was used to map the Illumina HiSeq and MinION reads to both contigs using the “map to reference” command with “medium sensitivity” settings. The smaller (1,400-bp) contig was present in the Illumina reads but not in the Nanopore reads. A BLASTn search of the NCBI
The gene prediction pipeline identified the small contig as similar (>97% identity) to DNA sequences originating from several ungulates, a *Clostridium botulinum* genome, an *Actinoballoteichus* sp. genome, and a *Babesia bigemina* genome. This observation, and the Geneious mapping, suggested that this second contig was a contaminant, and it was removed from the final assembly.

The remaining circular contig was annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v4.11 (17). The annotated chromosome for *Turicibacter sanguinis* MOL36T was 2,999,687 bp long with 34.4% G+C content. *Turicibacter sanguinis* MOL36T has 2,752 protein-coding sequences, with 43 pseudogenes and 121 tRNA genes.

Data availability. This genome sequence has been deposited in GenBank under the accession number CP053187. The Nanopore reads are available in the NCBI SRA under accession number SRR11185522.

ACKNOWLEDGMENTS

MiniION sequencing was performed at the NADC in Ames, Iowa. We thank Markus Göker (DSMZ) for permitting the use of his Illumina sequencing reads and Stefan Spring (DSMZ) for curating *T. sanguinis* MOL36T. We thank David Alt (USDA) for sequencing support.

This research was supported by USDA Agricultural Research Service (USDA-ARS) funds. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA.

REFERENCES