Complete Genome Sequence of *Streptococcus pneumoniae* Strain ST556, a Multidrug-Resistant Isolate from an Otitis Media Patient

Guiling Li, Fen Z. Hu, Xiaoan Wang, Yujun Cui, Jun Yang, Fen Qu, George F. Gao, and Jing-Ren Zhang

Center for Infectious Disease Research, School of Medicine, Tsinghua University, Beijing, China; Center for Genomic Sciences, Allegheny-Singer Research Institute, Pittsburgh, Pennsylvania, USA; State Key Laboratory of Pathogens and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing, China; Center for Immunology and Microbial Disease, Albany Medical College, Albany, New York, USA; The Center of Clinical Diagnosis, 302 Hospital of PLA, Beijing, China; and CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China

*Streptococcus pneumoniae* is a major pathogen causing bacterial infection in the middle ear of humans. We previously used *S. pneumoniae* strain ST556, a low-passage 19F isolate from an otitis media patient, to perform a whole-genome screen for ear infection-associated genes in a chinchilla model. This report presents the complete genome sequence of ST556. The genome sequence will provide information complementary to the experimental data from our genetic study of this strain.

*S. pneumoniae* is a nasopharyngeal commensal and also causes infections, including ear infection (or otitis media), pneumonia, and meningitis (10). In the context of high genome plasticity (3), this bacterium is classified into at least 92 serotypes based on the antigenic diversity in the capsular polysaccharide (6, 10, 12). *S. pneumoniae* strains of serotypes 14, 19, and 24 are among the most commonly encountered in otitis media patients (1, 7). We previously conducted the first whole-genome mutagenesis study to identify *S. pneumoniae* genes that are associated with bacterial infectivity (survival and growth) in the middle ear by using strain ST556 (2). ST556 is a low-passage multidrug-resistant serotype 19F isolate from an otitis media patient (7). That study led to the identification of 169 putative otitis media-associated *S. pneumoniae* genes (2). Due to the lack of genomic sequence information for this strain at the time, we reported our genetic data in the context of the gene number system of TIGR4, a capsular type 4 strain with a completely sequenced genome (13). To complement the *in vitro* screening data for ST556 (2), we have recently obtained the full genome sequence of this strain.

The genome was sequenced to 24.2-fold coverage using the 454 GS 20 sequencer as described previously (9). Lander-Waterman statistics predict that this level of coverage provided greater than 99.9% coverage of the genome. The Newbler *de novo* assembler used 200,678 reads with an average length of 250 bases to assemble the genome into 181 contigs as described previously (11). The 454-assembled contigs were ordered and oriented into scaffolds by alignment with complete *S. pneumoniae* genome sequences using Nucmer software (4) to identify the closest reference. The alignment included the genome sequences of strains R6 (accession no. NC_003098), D39 (accession no. NC_003098/NC_008533), TIGR4 (accession no. NC_003028), 670-6B (accession no. NC_014498), ATCC 700669 (accession no. NC_011900), Hungary19A-6 (accession no. NC_01380), TCH8431/19A (accession no. NC_014251), CGSP14 (accession no. NC_010582), Taiwan19F-14 (accession no. NC_012469), JHA (accession no. NC_012466), P1031 (accession no. NC_012467), 70858 (accession no. NC_012468), G54 (accession no. NC_011072), and AP200 (accession no. NC_014494). The sequence gaps were filled by PCR amplification and primer walking. Prediction of putative coding sequences and gene annotation were done by NCBI using the Prokaryotic Genomes Automatic Annotation Pipeline (http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html).

The ST556 genome consists of 2,145,902 nucleotides, with 2,162 predicted protein-encoding sequences and 4 rRNA loci. Among the strains with publicly available genome sequences, *S. pneumoniae* Taiwan19F-14 (accession no. NC_012469), a type 19F strain, is the closest to ST556 based on their similarities in genomic sequence and gene order, although the disease and isolation source associated with strain Taiwan19F-14 is unclear in the current genome annotation. A major difference between the genomes of ST556 and Taiwan19F-14 is the presence of an MM1-like phage (38,165 nucleotides) in the ST556 genome (5). The MM1 phage was previously reported to affect the colony opacity and adherence phenotypes of *S. pneumoniae* (8).

**Nucleotide sequence accession number.** The sequence of the *S. pneumoniae* ST556 genome has been deposited in GenBank under the accession number CP000357.

**ACKNOWLEDGMENTS**

We are grateful to M. R. Jacobs for providing the ST556 strain and H. Tettelin for valuable advice on genome annotation.

This work was supported by a grant (to J.-R.Z.) from the NIH/NIDCD (no. DC006917) and a grant (to J.-R.Z. and G.F.G.) from the China Natural Sciences Foundation (NSFC; no. 30728014).

**REFERENCES**


