



# Complete Genome Sequence of *Mycoplasma pneumoniae* Type 2 Reference Strain FH Using Single-Molecule Real-Time Sequencing Technology

Heta P. Desai, Shatavia S. Morrison, Maureen H. Diaz, Alvaro J. Benitez,  
Bernard J. Wolff, Jonas M. Winchell

Division of Bacterial Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

**ABSTRACT** *Mycoplasma pneumoniae* type 2 strain FH was previously sequenced with Illumina (FH-Illumina) and 454 (FH-454) technologies according to Xiao et al. (2015) and Krishnakumar et al. (2010). Comparative analyses revealed differences in genomic content between these sequences, including a 6-kb region absent from the FH-454 submission. Here, we present a complete genome sequence of FH sequenced with the Pacific Biosciences RSII platform.

*Mycoplasma pneumoniae* is a common cause of community-acquired pneumonia worldwide (1). Two primary strain types, type 1 and type 2, are recognized based on variations within the genes encoding the P1 adhesion molecule. Complete genomes of the reference strains M129 (type 1) and FH (type 2) have been available since 1996 and 2010, respectively (2–5). Comparative genomics relies on complete, curated references, as even minor differences may impact transmissibility or virulence. Single-molecule real-time (SMRT) sequencing generates long reads for assembly of genomes with improved quality, particularly across difficult-to-sequence regions. Here, we present a complete sequence of strain FH (FH-PacBio) using SMRT technology.

Genomic DNA was extracted with the MasterPure Complete DNA/RNA purification kit (Epicentre, Madison, WI, USA), and a 10-kb library was prepared for PacBio sequencing (Pacific Biosciences, Menlo Park, CA, USA). A paired-end Illumina sequencing library was also prepared using the NEBNext Ultra DNA kit (New England Biolabs, Ipswich, MA, USA). A total of 38,894 PacBio sequencing reads were assembled using the Hierarchical Genome Assembly Process (HGAP version 3) into a completely closed genome. Nucleotide accuracy was >99.9%, as indicated by the mapping of Illumina reads ( $n = 2,729,324$ ) to the assembled genome. The genome size was 817,276 bp. Prokka version 1.8 was used to predict 757 coding sequences, 36 tRNAs, and one tmRNA (6). In a pairwise comparison, the FH-Illumina sequence (NZ\_CP010546.1) was 69 bp shorter than FH-PacBio due to differences in a short tandem repeat region, and had only one fewer gene (7, 8). The largest discrepancy was found in comparing FH-454 (NC\_017504.1) and FH-PacBio; a 6-kb region containing lipoprotein-encoding genes was absent in FH-454, confirming a previous observation in comparing FH-Illumina and FH-454 (9). This difference may be due to misassembly resulting from a neighboring repetitive element or a gene knockout experiment performed on the isolate prior to sequencing (2). Prior to the availability of the FH-Illumina and FH-PacBio sequences, this 6-kb difference was thought to be a differentiating feature between variants of type 2 *M. pneumoniae* isolates. However, it is now recognized as a common feature to all type 2 strains. FH-PacBio and FH-Illumina, which differed only by 24 single nucleotide

Received 2 December 2016 Accepted 16 December 2016 Published 23 February 2017

**Citation** Desai HP, Morrison SS, Diaz MH, Benitez AJ, Wolff BJ, Winchell JM. 2017. Complete genome sequence of *Mycoplasma pneumoniae* type 2 reference strain FH using single-molecule real-time sequencing technology. *Genome Announc* 5:e01629-16. <https://doi.org/10.1128/genomeA.01629-16>.

**Copyright** © 2017 Desai 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 Jonas M. Winchell, [jwinchell@cdc.gov](mailto:jwinchell@cdc.gov).

polymorphisms, are more accurate reference genomes for FH and serve as superior representatives of type 2 *M. pneumoniae* genomes.

This analysis demonstrates the impact of sequencing technology and assembly methods on the generation of completely characterized reference genomes, which are necessary for understanding strain relatedness and identifying relevant features that may impact disease. The availability of well-curated genomes is imperative as whole-genome sequencing technology expands in public health and clinical laboratories.

**Accession number(s).** This whole-genome sequence has been deposited at GenBank under accession number [CP017327](#) and at the Sequence Read Archive under accession number SRR3924617.

## ACKNOWLEDGMENTS

The findings and conclusions in this presentation are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

## REFERENCES

1. Diaz MH, Benitez AJ, Winchell JM. 2015. Investigations of *Mycoplasma pneumoniae* infections in the United States: trends in molecular typing and macrolide resistance from 2006 to 2013. *J Clin Microbiol* 53:124–130. <https://doi.org/10.1128/JCM.02597-14>.
2. Krishnakumar R, Assad-Garcia N, Benders GA, Phan Q, Montague MG, Glass JI. 2010. Targeted chromosomal knockouts in *Mycoplasma pneumoniae*. *Appl Environ Microbiol* 76:5297–5299. <https://doi.org/10.1128/AEM.00024-10>.
3. Himmelreich R, Hilbert H, Plagens H, Pirkel E, Li BC, Herrmann R. 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res* 24:4420–4449. <https://doi.org/10.1093/nar/24.22.4420>.
4. Lluch-Senar M, Cozzuto L, Cano J, Delgado J, Llórens-Rico V, Pereyre S, Bébear C, Serrano L. 2015. Comparative “-omics” in *Mycoplasma pneumoniae* clinical isolates reveals key virulence factors. *PLoS One* 10: e0137354. <https://doi.org/10.1371/journal.pone.0137354>.
5. Touati A, Blouin Y, Sirand-Pugnet P, Renaudin H, Oishi T, Vergnaud G, Bébear C, Pereyre S. 2015. Molecular epidemiology of *Mycoplasma pneumoniae*: genotyping using single nucleotide polymorphisms and SNaPshot technology. *J Clin Microbiol* 53:3182–3194. <https://doi.org/10.1128/JCM.01156-15>.
6. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
7. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14:1394–1403. <https://doi.org/10.1101/gr.2289704>.
8. Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. <https://doi.org/10.1371/journal.pone.0011147>.
9. Xiao L, Ptacek T, Osborne JD, Crabb DM, Simmons WL, Lefkowitz EJ, Waites KB, Atkinson TP, Dybvig K. 2015. Comparative genome analysis of *Mycoplasma pneumoniae*. *BMC Genomics* 16:610. <https://doi.org/10.1186/s12864-015-1801-0>.