

Whole-Genome Sequence of the Human Pathogen *Legionella pneumophila* Serogroup 12 Strain 570-CO-H

Francisco Amaro,^a Jack A. Gilbert,^{b,c} Sarah Owens,^b William Trimble,^b and Howard A. Shuman^a

Department of Microbiology, University of Chicago, Chicago, Illinois, USA^a; Institute for Genomic and Systems Biology, Argonne National Laboratory, Lemont, Illinois, USA^b; and Department of Ecology and Evolution, University of Chicago, Chicago, Illinois, USA^c

We present the genomic sequence of the human pathogen *Legionella pneumophila* serogroup 12 strain 570-CO-H (ATCC 43290), a clinical isolate from the Colorado Department of Health, Denver, CO. This is the first example of a genome sequence of *L. pneumophila* from a serogroup other than serogroup 1. We highlight the similarities and differences relative to six genome sequences that have been reported for serogroup 1 strains.

The Gram-negative gammaproteobacterium *Legionella pneumophila* is an opportunistic human pathogen that causes Legionnaires' disease, an acute pneumonia acquired from aerosolized contaminated fresh water. Currently only the genomes of *L. pneumophila* serogroup 1 strains—130b (10), Alcoy (4), Corby (7), Lens (1), Paris (1), Philadelphia 1 (2)—have been reported. Whole-genome analysis revealed a highly plastic genome and a large repertoire of effector proteins, which may be a consequence of the ability of legionellae to grow within a wide range of phylogenetically diverse protozoan hosts and human macrophages (6).

The *L. pneumophila* serogroup 12 strain 570-CO-H (ATCC 43290), whose genome was sequenced here, is a clinical isolate from the Colorado Department of Health, Denver, CO (human isolate 570-CO-H) (11). For whole-genome sequencing of strain 570-CO-H, short-insert and long-insert shotgun libraries were generated from 5 μ g of genomic DNA using standard Illumina protocols. A total of 7,061,382 (416-fold coverage) and 3,023,590 (178-fold coverage) reads for each library were generated using an Illumina DNA analyzer and *de novo* assembled into 892 contigs larger than 900 bp with Geneious software. Contigs generated from both libraries were further analyzed and assembled into one scaffold with Geneious.

Open reading frames predicted by Glimmer 3 (5) were annotated by BLAST searches against the GenBank nonredundant nucleotide and protein databases. tRNA genes were identified with the tRNAscanSE program (9), and rRNA genes were identified by a BLAST search. Gene names and protein products were assigned based on significant homology to known genes found by BLAST and homology searches against previously sequenced genomes of *L. pneumophila* strains. Mauve (3) was used to compare the genome of 570-CO-H with the other sequenced *L. pneumophila* genomes (strains Alcoy, Corby, Lens, Paris, Philadelphia, and 130b).

The complete genome of *L. pneumophila* strain 570-CO-H consists of a single circular chromosome of approximately 3.36 Mb with an average G+C content of 38.2%. The genome sequence is predicted to contain 3 rRNA operons, 43 tRNA genes, and 2,942 putative coding sequences, of which 837 are for hypothetical proteins (28%). Whole-genome comparative analysis showed that the genome of 570-CO-H is highly syntenic with those of the six sequenced *L. pneumophila* serogroup 1 strains, although duplications, deletions, and translocations of gene clusters were detected. Genes encoding all the structural components of the type I secretion system (T1SS), the T2SS, and the Icm/Dot T4BSS were found

in syntenic positions. However, the gene cluster encoding the Lvh T4ASS was not present in the 570-CO-H genome. Of the proven 275 Icm/Dot translocated effectors (8), 274 were identified in the 570-CO-H genome. The proposed core set of 220 effector genes common to all six sequenced serogroup 1 strains (8) was also present in the 570-CO-H genome. A total of 117 genes encoding eukaryotic-type proteins present in other *L. pneumophila* strains were also identified. Major differences from other *L. pneumophila* genomes include a particular 46-kb cluster of genes encoding proteins involved in lipopolysaccharide biosynthesis and a cluster of seven hypothetical proteins that are not present in other *L. pneumophila* genomes.

Nucleotide sequence accession number. The *L. pneumophila* serogroup 12 strain 570-CO-H whole-genome sequence has been deposited at GenBank under accession number [CP003192](https://www.ncbi.nlm.nih.gov/nuccore/CP003192).

ACKNOWLEDGMENTS

This work was supported by NIAID RO1 AI23549 (H.A.S.) and the U.S. Department of Energy under contract DE-AC02-06CH11357 (J.A.G.). F. Amaro is the recipient of a postdoctoral fellowship from the Fulbright Commission and Ministry of Education of Spain.

REFERENCES

1. Cazalet C, et al. 2004. Evidence in the *Legionella pneumophila* genome for exploitation of host cell functions and high genome plasticity. *Nat. Genet.* 36:1165–1173.
2. Chien M, et al. 2004. The genomic sequence of the accidental pathogen *Legionella pneumophila*. *Science* 305:1966–1968.
3. Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14:1394–1403.
4. D'Auria G, Jimenez-Hernandez N, Peris-Bondia F, Moya A, Latorre A. 2010. *Legionella pneumophila* pangenome reveals strain-specific virulence factors. *BMC Genomics* 11:181.
5. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679.
6. Fields BS. 1996. The molecular ecology of legionellae. *Trends Microbiol.* 4:286–290.

Received 14 December 2011 Accepted 4 January 2012

Address correspondence to Howard A. Shuman, hshuman@bsd.uchicago.edu.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.06626-11

7. Glockner G, et al. 2008. Identification and characterization of a new conjugation/type IVA secretion system (trb/tra) of *Legionella pneumophila* Corby localized on two mobile genomic islands. *Int. J. Med. Microbiol.* 298:411–428.
8. Gomez-Valero L, Rusniok C, Cazalet C, Buchrieser C. 2011. Comparative and functional genomics of *Legionella* identified eukaryotic like proteins as key players in host-pathogen interactions. *Front. Microbiol.* 2:208.
9. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25:955–964.
10. Schroeder GN, et al. 2010. *Legionella pneumophila* strain 130b possesses a unique combination of type IV secretion systems and novel Dot/Icm secretion system effector proteins. *J. Bacteriol.* 192:6001–6016.
11. Thacker WL, Wilkinson HW, Benson RF, Brenner DJ. 1987. *Legionella pneumophila* serogroup 12 isolated from human and environmental sources. *J. Clin. Microbiol.* 25:569–570.