

# Complete Genome Sequence of *Haemophilus influenzae* Strain 375 from the Middle Ear of a Pediatric Patient with Otitis Media

Joshua Chang Mell,<sup>a,b,c</sup> Sunita Sinha,<sup>d</sup> Sergey Balashov,<sup>a,e</sup> Cristina Viadas,<sup>f,g</sup> Christopher J. Grassa,<sup>b,h</sup> Garth D. Ehrlich,<sup>a,e</sup> Corey Nislow,<sup>d</sup> Rosemary J. Redfield,<sup>b,c</sup> Junkal Garmendia<sup>f,g</sup>

Centers for Genomic Sciences and Advanced Microbial Processing, Institute for Molecular Medicine & Infectious Disease, Department of Microbiology and Immunology, Drexel University College of Medicine, Philadelphia, Pennsylvania, USA<sup>a</sup>; Biodiversity Research Centre, University of British Columbia, Vancouver, BC, Canada<sup>b</sup>; Department of Zoology, University of British Columbia, Vancouver, BC, Canada<sup>c</sup>; Department of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada<sup>d</sup>; Genomic Core Facility, Institute for Clinical and Translational Research, Drexel University College of Medicine, Philadelphia, Pennsylvania, USA<sup>e</sup>; Instituto de Agrobiotecnología, CSIC-Universidad Pública Navarra-Gobierno, Navarra, Spain<sup>f</sup>; CIBERES, Madrid, Spain<sup>g</sup>; Department of Botany, University of British Columbia, Vancouver, BC, Canada<sup>h</sup>

**Originally isolated from a pediatric patient with otitis media, *Haemophilus influenzae* strain 375 (Hi375) has been extensively studied as a model system for intracellular invasion of airway epithelial cells and other pathogenesis traits. Here, we report its complete genome sequence and methylome.**

Received 18 October 2014 Accepted 21 October 2014 Published 4 December 2014

**Citation** Mell JC, Sinha S, Balashov S, Viadas C, Grassa CJ, Ehrlich GD, Nislow C, Redfield RJ, Garmendia J. 2014. Complete genome sequence of *Haemophilus influenzae* strain 375 from the middle ear of a pediatric patient with otitis media. *Genome Announc.* 2(6):e01245-14. doi:10.1128/genomeA.01245-14.

**Copyright** © 2014 Mell et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Joshua Chang Mell, [joshua.mell@drexelmed.edu](mailto:joshua.mell@drexelmed.edu).

*Haemophilus influenzae* is a diverse bacterium, usually associated with human nasopharyngeal carriage but it can also be a potent pathogen. Although an effective vaccine against meningitis-causing type b strains is in wide use, nontypeable *H. influenzae* (NTHi) remains a common problem in patients with chronic respiratory conditions and pediatric ear infections (1, 2). The NTHi otitis media isolate Hi375 has been extensively used in studies of bacterial pathogenesis, particularly with respect to intracellular invasion of airway epithelia, outer membrane physiology, and animal models of pathogenesis (3–7).

Genomic DNA was extracted by the CTAB method (8), and sequencing libraries were constructed according to the manufacturers' instructions using Nextera XT for Illumina and the 6-kb insert protocol for PacBio. Illumina sequencing was part of multiplexed HiSeq RapidRuns, and  $\sim 3.8 \times 10^7$  read pairs ( $2 \times 101$  nt) were collected for Hi375 ( $\sim 4,000$ -fold coverage). PacBio sequencing (v 2.1.0) was performed using a single SMRTcell with P4-C2 chemistry. A 2-h movie generated 44,007 polymerase reads ( $N_{50} = 5,022$  nucleotide (nt); postfiltered subreads,  $N_{50} = 3,116$  nt).

*De novo* assembly of Illumina reads trimmed adapters with Trimmomatic (9), merged overlapping reads with COPE (10), and assembled with RAY (11), as previously described (12), yielding 21 contigs. This assembly was reconciled using CISA (13) with another partial assembly of Hi375 (14), producing a merged assembly of 16 contigs covering 1,824,471 bp.

*De novo* assembly of PacBio data with the HGAP assembler (15) (v3beta) yielded a single contig with mean coverage of 66-fold (1 < 3-kb contig with coverage < 10-fold was discarded). Circular closure used Minimus2 (<http://amos.sourceforge.net/wiki/index.php/Minimus2>) to trim the ends and permute the genome to begin at the DnaA gene (identified by BLAST), followed by Quiver-based error correction (15) for a final closed genome size of 1,850,897 bp. Assembly accuracy was verified using

Mauve (16) to reorder Illumina contigs against the complete assembly, finding perfect synteny. Illumina read pairs were aligned to the complete assembly using bwa mem (17) and sambamba (<https://github.com/lomereiter/sambamba>). Subsequently samtools mpileup and bcftools view (18) identified no variants with quality of >30.

The Pacific Biosciences “Modification and Motif Analysis” pipeline (v1) identified six 6-methyladenine motifs (bold underlined positions at Ts indicating methylation on the reverse complement): GATC, CCGAA, GACCN<sub>6</sub>GTT, ATGN<sub>6</sub>CCT, TCAN<sub>6</sub>TRCC, AACN<sub>6</sub>RTC. Additionally, an unknown cytosine modification motif was identified (GCGCGCBHV).

Results with a streptomycin-resistant (Str<sup>r</sup>) derivative—created by transformation with a PCR fragment from a multidrug resistant Rd derivative, MAP7 (coordinates 599,059 to 602,433 of the Rd genome, NC\_000907.1) were comparable to those described above for Hi375. A single circularized contig was generated, and short reads agreed with the assembly. Eight single-nucleotide variants distinguished this strain from Hi375. As expected, all were clustered at *rpsL* (30S ribosomal protein S12), including the Str<sup>r</sup> allele, an A → G transition at position 444,369. The remaining variants were the next seven that distinguish Hi375 from Rd.

Annotation by the NCBI prokaryotic genome annotation pipeline found the Hi375 chromosome contains 1,699 coding sequences, 6 rRNA clusters, and 59 tRNAs, covering all 20 amino acids including selenocysteine. We expect this complete genome to facilitate molecular genomics investigations into NTHi pathogenesis.

**Nucleotide sequence accession number.** The complete genome of nontypeable *Haemophilus influenzae* strain 375 was submitted to NCBI under the accession number CP009610. This is the first version of the complete sequence.

## ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health Ruth Kirschstein Postdoctoral Fellowship to J.C.M. and R01 DC0214 to G.D.E., a Canadian Institutes of Health Research grant to R.J.R., and MINECO SAF2012-31166 and CIBERES funding to J.G.

Illumina sequencing was performed at the Pharmaceutical Sciences Sequencing Centre at the University of British Columbia, and Pacific Biosciences sequencing was performed at the Genomics Core Facility in the Institute for Clinical and Translational Research at the Drexel University College of Medicine.

## REFERENCES

- Clementi CF, Murphy TF. 2011. Non-typeable *Haemophilus influenzae* invasion and persistence in the human respiratory tract. *Front. Cell. Infect. Microbiol.* 1:1. <http://dx.doi.org/10.3389/fcimb.2011.00001>.
- Jalalvand F, Riesbeck K. 2014. *Haemophilus influenzae*: recent advances in the understanding of molecular pathogenesis and polymicrobial infections. *Curr. Opin. Infect. Dis.* 27:268–274. <http://dx.doi.org/10.1097/QCO.0000000000000056>.
- Hood DW, Makepeace K, Deadman ME, Rest RF, Thibault P, Martin A, Richards JC, Moxon ER. 1999. Sialic acid in the lipopolysaccharide of *Haemophilus influenzae*: strain distribution, influence on serum resistance and structural characterization. *Mol. Microbiol.* 33:679–692. <http://dx.doi.org/10.1046/j.1365-2958.1999.01509.x>.
- Bouchet V, Hood DW, Li J, Brisson JR, Randle GA, Martin A, Li Z, Goldstein R, Schweda EK, Pelton SI, Richards JC, Moxon ER. 2003. Host-derived sialic acid is incorporated into *Haemophilus influenzae* lipopolysaccharide and is a major virulence factor in experimental otitis media. *Proc. Natl. Acad. Sci. U. S. A.* 100:8898–8903. <http://dx.doi.org/10.1073/pnas.1432026100>.
- Morey P, Cano V, Martí-Llitas P, López-Gómez A, Regueiro V, Saus C, Bengoechea JA, Garmendia J. 2011. Evidence for a non-replicative intracellular stage of nontypable *Haemophilus influenzae* in epithelial cells. *Microbiology* 157:234–250. <http://dx.doi.org/10.1099/mic.0.040451-0>.
- López-Gómez A, Cano V, Moranta D, Morey P, García del Portillo F, Bengoechea JA, Garmendia J. 2012. Host cell kinases,  $\alpha 5$  and  $\beta 1$  integrins, and Rac1 signalling on the microtubule cytoskeleton are important for non-typable *Haemophilus influenzae* invasion of respiratory epithelial cells. *Microbiology* 158:2384–2398. <http://dx.doi.org/10.1099/mic.0.059972-0>.
- Morey P, Viadas C, Euba B, Hood DW, Barberán M, Gil C, Grilló MJ, Bengoechea JA, Garmendia J. 2013. Relative contributions of lipooligosaccharide inner and outer core modifications to nontypable *Haemophilus influenzae* pathogenesis. *Infect. Immun.* 81:4100–4111. <http://dx.doi.org/10.1128/IAI.00492-13>.
- Wilson K. 2001. Preparation of genomic DNA from bacteria. *Curr. Protoc. Mol. Biol.* Chapter 2:Unit 2.4. <http://dx.doi.org/10.1002/0471142727.mb0204s56>.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <http://dx.doi.org/10.1093/bioinformatics/btu170>.
- Liu B, Yuan J, Yiu SM, Li Z, Xie Y, Chen Y, Shi Y, Zhang H, Li Y, Lam TW, Luo R. 2012. COPE: an accurate k-mer-based pair-end reads connection tool to facilitate genome assembly. *Bioinformatics* 28:2870–2874. <http://dx.doi.org/10.1093/bioinformatics/bts563>.
- Boisvert S, Laviolette F, Corbeil J. 2010. Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J. Comput. Biol.* 17:1519–1533. <http://dx.doi.org/10.1089/cmb.2009.0238>.
- Garmendia J, Viadas C, Calatayud L, Mell JC, Martí-Llitas P, Euba B, Llobet E, Gil C, Bengoechea JA, Redfield RJ, Linares J. 2014. Characterization of nontypable *Haemophilus influenzae* isolates recovered from adult patients with underlying chronic lung disease reveals genotypic and phenotypic traits associated with persistent infection. *PLoS One* 9:e97020. <http://dx.doi.org/10.1371/journal.pone.0097020>.
- Lin SH, Liao YC. 2013. CISA: contig integrator for sequence assembly of bacterial genomes. *PLoS One* 8:e60843. <http://dx.doi.org/10.1371/journal.pone.0060843>.
- De Chiara M, Hood D, Muzzi A, Pickard DJ, Perkins T, Pizza M, Dougan G, Rappuoli R, Moxon ER, Soriani M, Donati C. 2014. Genome sequencing of disease and carriage isolates of nontypable *Haemophilus influenzae* identifies discrete population structure. *Proc. Natl. Acad. Sci. U. S. A.* 111:5439–5444. <http://dx.doi.org/10.1073/pnas.1403353111>.
- Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10:563–569. <http://dx.doi.org/10.1038/nmeth.2474>.
- Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. <http://dx.doi.org/10.1371/journal.pone.0011147>.
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997*. <http://arxiv.org/abs/1303.3997>.
- Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27:2987–2993. <http://dx.doi.org/10.1093/bioinformatics/btr509>.