

## Genome Sequences for Five Strains of the Emerging Pathogen *Haemophilus haemolyticus*

I. King Jordan,<sup>1,2,3</sup> Andrew B. Conley,<sup>2</sup> Ivan V. Antonov,<sup>2</sup> Robert A. Arthur,<sup>2</sup> Erin D. Cook,<sup>2</sup> Guy P. Cooper,<sup>2</sup> Bernard L. Jones,<sup>2</sup> Kristen M. Knipe,<sup>2</sup> Kevin J. Lee,<sup>2</sup> Xing Liu,<sup>2</sup> Gabriel J. Mitchell,<sup>2</sup> Pushkar R. Pande,<sup>2</sup> Robert A. Petit,<sup>2</sup> Shaopu Qin,<sup>2</sup> Vani N. Rajan,<sup>2</sup> Shruti Sarda,<sup>2</sup> Aswathy Sebastian,<sup>2</sup> Shiyuyun Tang,<sup>2</sup> Racchit Thapliyal,<sup>2</sup> Neha J. Varghese,<sup>2</sup> Tianjun Ye,<sup>2</sup> Lee S. Katz,<sup>4</sup> Xin Wang,<sup>4</sup> Lori Rowe,<sup>5</sup> Michael Frace,<sup>5</sup> and Leonard W. Mayer<sup>4\*</sup>

School of Biology, Georgia Institute of Technology, Atlanta, Georgia 30332<sup>1</sup>; Computational Genomics Group, Georgia Institute of Technology, Atlanta, Georgia 30332<sup>2</sup>; PanAmerican Bioinformatics Institute, Santa Marta, Magdalena, Colombia<sup>3</sup>; Meningitis and Vaccine Preventable Diseases Branch, Centers for Disease Control and Prevention, Atlanta, Georgia 30333<sup>4</sup>; and Biotechnology Core Facility Branch, Division of Scientific Resources, National Center for Preparedness, Detection, and Control of Infectious Diseases (NCPDCID), CDC, Atlanta, Georgia 30333<sup>5</sup>

Received 21 July 2011/Accepted 1 August 2011

**We report the first whole-genome sequences for five strains, two carried and three pathogenic, of the emerging pathogen *Haemophilus haemolyticus*. Preliminary analyses indicate that these genome sequences encode markers that distinguish *H. haemolyticus* from its closest *Haemophilus* relatives and provide clues to the identity of its virulence factors.**

The bacterium *Haemophilus haemolyticus* is a Gram-negative, facultative anaerobe that colonizes the human respiratory tract (10). *H. haemolyticus* is one of eight *Haemophilus* species and is a sister taxon to the most pathogenic member of the genus, *H. influenzae*. Until very recently, *H. haemolyticus* has been regarded as a strict human commensal that rarely causes invasive disease, such as meningitis and bacteremia (1, 4, 9; R. D. Mair, X. Wang, E. Briere, L. S. Katz, A. Cohn, and L. W. Mayer, unpublished data). However, five cases of invasive disease detected in 2009 and 2010 which were originally attributed to nontypeable *H. influenzae* were later confirmed to be caused by *H. haemolyticus* (Mair et al., unpublished). The nature of its virulence determinants is unknown, and unambiguously differentiating *H. haemolyticus* from *H. influenzae* is challenging due to their close genetic relatedness. Genome sequence analysis may provide important clues regarding these open questions. Here, we report genome sequences for five strains of *H. haemolyticus*; these are the first reported whole-genome sequences for this species.

Two carriage *H. haemolyticus* strains were taken from a 2009 carriage study in Minnesota (M19107 and M19501) (6), and three pathogenic strains were isolated from patients in Georgia (M21127), Texas (M21621), and Illinois (M21639) in 2010. Sequencing of these five *H. haemolyticus* strains was performed at the Centers for Disease Control and Prevention (CDC) Biotechnology Core Facility using Roche Applied Science 454

pyrosequencing with the GS FLX titanium platform. The genomes were sequenced at 13.2× to 69.3× coverage (average, 37.8×), and the predicted genome sizes ranged from 1.89 to 2.33 Mb (average, 2.03 Mb). Genome assembly and annotation were performed by the Computational Genomics Group at the Georgia Institute of Technology using a modified version of the CG-Pipeline annotation platform (freely available at <http://sourceforge.net/projects/cg-pipeline/>) (5). For each strain, independent genome assemblies were first constructed using the Newbler (7) and Mira (3) assemblers and then merged using Minimus (12); genes were predicted using *ab initio* prediction with GeneMarkS (2). Contig numbers ranged from 22 to 123 (average, 53), and there were 1,886 to 2,782 predicted genes per genome (average, 2,190).

*H. haemolyticus* is most closely related to *H. influenzae*, and there is currently no molecular typing scheme that can unambiguously distinguish the two species (8, 11). For example, neither 16S rRNA gene sequences nor multilocus sequence typing can be used to unambiguously delineate *H. haemolyticus* from nontypeable *H. influenzae* strains. This provides a fundamental challenge to surveillance, and indeed, invasive cases of *H. haemolyticus* have been erroneously attributed to *H. influenzae* (Mair et al., unpublished). Complete genome sequences can provide a source of novel genetic markers that may be able to distinguish between these closely related species. Preliminary comparative analysis of the five genome sequences of *H. haemolyticus* characterized here with 19 complete *H. influenzae* genome sequences revealed 54 clusters of orthologous genes that are present exclusively among *H. haemolyticus* bacteria and absent from *H. influenzae* bacteria. These gene sequences provide potential markers that can be used as the basis of

\* Corresponding author. Mailing address: Meningitis and Vaccine Preventable Diseases Branch, Centers for Disease Control and Prevention, Atlanta, GA 30333. Phone: (404) 639-2841. Fax: (404) 639-4421. E-mail: lwm1@cdc.gov.

future molecular typing assays, and the presence of such lineage-specific genes also suggests the possibility of a specific genomic basis of virulence for pathogenic *H. haemolyticus* strains.

**Nucleotide sequence accession numbers.** The five *H. haemolyticus* whole-genome sequence assemblies and their annotations were deposited in GenBank under the accession numbers AFQN00000000 (M19107), AFQO00000000 (M19501), AFQP00000000 (M21127), AFQQ00000000 (M21621), and AFQR00000000 (M21639).

#### ACKNOWLEDGMENTS

This work was supported by an Alfred P. Sloan Research Fellowship in Computational and Evolutionary Molecular Biology (grant BR-4839 to I.K.J.) and by the Georgia Research Alliance (GRA.VAC09.O to I.K.J. and L.W.M.).

We acknowledge the support of the Georgia Tech graduate programs in bioinformatics. We thank the Active Bacterial Core surveillance team, Roman Golash from the Illinois Department of Health, and the Texas Department of State Health Services for providing strains.

#### REFERENCES

1. Albritton, W. L. 1982. Infections due to *Haemophilus* species other than *H. influenzae*. *Annu. Rev. Microbiol.* **36**:199–216.
2. Besemer, J., A. Lomsadze, and M. Borodovsky. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* **29**:2607–2618.
3. Chevreur, B., et al. 2004. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res.* **14**:1147–1159.
4. De Santo, D. A., and M. White. 1933. *Haemophilus haemolyticus* Endocarditis. *Am. J. Pathol.* **9**:381–392.
5. Kislyuk, A. O., et al. 2010. A computational genomics pipeline for prokaryotic sequencing projects. *Bioinformatics* **26**:1819–1826.
6. Lowther, S. A., et al. 18 May 2011, posting date. *Haemophilus influenzae* type b infection, vaccination, and *H. influenzae* carriage in children in Minnesota, 2008–2009. *Epidemiol. Infect.* [Epub ahead of print.] doi:10.1017/S0950268811000793.
7. Margulies, M., et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**:376–380.
8. McCrean, K. W., et al. 2008. Relationships of nontypeable *Haemophilus influenzae* strains to hemolytic and nonhemolytic *Haemophilus haemolyticus* strains. *J. Clin. Microbiol.* **46**:406–416.
9. Miller, C. P., and A. Branch. 1923. Subacute bacterial endocarditis due to a hemolytic hemophilic bacillus. *Arch. Intern. Med.* **32**:911–926.
10. Mukundan, D., Z. Ecevit, M. Patel, C. F. Marrs, and J. R. Gilsdorf. 2007. Pharyngeal colonization dynamics of *Haemophilus influenzae* and *Haemophilus haemolyticus* in healthy adult carriers. *J. Clin. Microbiol.* **45**:3207–3217.
11. Nørskov-Lauritsen, N., M. D. Overballe, and M. Kilian. 2009. Delineation of the species *Haemophilus influenzae* by phenotype, multilocus sequence phylogeny, and detection of marker genes. *J. Bacteriol.* **191**:822–831.
12. Sommer, D. D., A. L. Delcher, S. L. Salzberg, and M. Pop. 2007. Minimus: a fast, lightweight genome assembler. *BMC Bioinformatics* **8**:64.