AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Draft Genome Sequences for a Diverse Set of Seven *Haemophilus* and *Aggregatibacter* Species

Megan Nichols,[a] Nadav Topaz,[b] Xiong Wang,[a,c] Xin Wang,[b] Dave Boxrud[a]

[a]Minnesota Department of Health Public Health Laboratory, St. Paul, Minnesota, USA

[b]Meningitis and Vaccine Preventable Diseases Branch, Division of Bacterial Diseases, National Center for Immunization and Respiratory Diseases, Coordinating Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

[c]Department of Veterinary Biomedical Sciences, University of Minnesota, St. Paul, Minnesota, USA

ABSTRACT *Haemophilus* is a complex genus that includes commensal and pathogenic species that pose a public health threat to humans. While the pathogenic species have been studied extensively, many commensals have limited genomic information available. Here, we present 24 draft genomes for a diverse set of 7 *Haemophilus* and *Aggregatibacter* species.

The genus *Haemophilus* consists of pleomorphic Gram-negative coccobacilli that all share similar growth requirements for the presence of hemin and/or NAD (1). *Haemophilus* species are part of the commensal flora in humans and can most commonly be found colonizing the upper respiratory tract, oral cavity, and mucosal membranes (1). The human pathogen belonging to the genus is *Haemophilus influenzae*, which can cause a variety of conditions, including meningitis, bacteremia, otitis media, sinusitis, and conjunctivitis (2). While the pathogenic species have been studied extensively, other members of the genus have been studied far less extensively, and many have only limited available genomic information. To supplement the existing *Haemophilus* genomic collection, we present genomic data for 7 *Haemophilus* and *Aggregatibacter* species, *Aggregatibacter aphrophilus* (formerly *H. aphrophilus*), *Aggregatibacter segnis* (formerly *H. segnis*), *H. haemolyticus*, *H. parahaemolyticus*, *H. parainfluenzae*, *H. paraphrohaemolyticus*, and *H. sputorum*.

Bacteria were isolated from clinical specimens collected in Minnesota from 2000 to 2015, and single colonies were cultivated on chocolate agar for 24 to 48 h at 33 to 37°C and 4 to 6% CO_2. Bacterial DNA was extracted using the QIAmp DNA blood minikit on the Qiagen QIAcube following the manufacturer's guidelines, and DNA concentrations were quantitated using the Qubit double-stranded-DNA (dsDNA) high-sensitivity (HS) assay kit (Thermo Fisher Scientific). Samples were prepared for whole-genome sequencing following the Nextera XT DNA Library preparation protocol and the manufacturer's (Illumina) guidelines. Bar-coded libraries were then pooled and loaded onto the Illumina MiSeq system using 500-cycle V2 chemistries for multiplexed 250-bp paired-end sequencing. The Illumina reads were then trimmed using Cutadapt 1.8 (3) with default parameters and assembled using SPAdes 3.7.0 (4) with default parameters. Genomes were annotated by NCBI using the Prokaryotic Genome Annotation Pipeline (PGAP) (5).

The genus *Haemophilus* has undergone many revisions over the years; with the addition of new species to the genus (*H. pittmaniae* and *H. sputorum*) and reclassification of six former members, including two species covered in this study (*A. aphrophilus* and *A. segnis*), the taxonomy of the genus is an ongoing topic of discussion (6). Despite the importance of rapid and accurate species identification in clinical and research settings, correct identification of *Haemophilus* has been a continuous challenge due to the lack of proper detection methods. Not all species are clearly distinguishable by their

Received 27 June 2018 **Accepted** 2 October 2018 **Published** 25 October 2018

Citation Nichols M, Topaz N, Wang X, Wang X, Boxrud D. 2018. Draft genome sequences for a diverse set of seven *Haemophilus* and *Aggregatibacter* species. Microbiol Resour Announc 7:e00880-18. https://doi.org/10.1128/MRA.00880-18.

Editor Jason Stajich, University of California, Riverside

Address correspondence to Dave Boxrud, dave.boxrud@state.mn.us.

TABLE 1 Metadata and GenBank accession numbers for draft genome assemblies reported in this study

Isolate name	GenBank accession no.	Species	Collection yr	Isolation source	N_{50} (bp)	No. of contigs	Avg coverage (×)
C2015005679	QEQH00000000	*H. sputorum*	2015	Blood	201,027	26	43
C2015005473	QEQG00000000	*H. sputorum*	2015	Blood	201,027	26	34
C2014016342	QEQF00000000	*H. paraphrohaemolyticus*	2014	Bronchus	246,469	27	27
C2011020591	QEQE00000000	*H. parainfluenzae*	2011	Sputum	254,703	26	37
C2010039593	QEQD00000000	*H. parahaemolyticus*	2010	Sputum	171,434	41	36
C2010020251	QEQC00000000	*A. aphrophilus*	2010	Hip	402,864	18	38
C2009038101	QEQB00000000	*H. parainfluenzae*	2009	Sputum	230,991	25	45
C2009017515	QEQA00000000	*A. aphrophilus*	2009	Cheek	399,989	35	32
C2008003258	QEPZ00000000	*H. parainfluenzae*	2008	Sputum	189,256	21	43
C2008003249	QEPY00000000	*A. aphrophilus*	2008	Sinus	451,174	18	25
C2008001782	QEPX00000000	*A. aphrophilus*	2008	Bronchial wash	347,044	16	33
C2008001710	QEPW00000000	*H. parainfluenzae*	2008	Sputum	112,625	51	61
C2008001229	QEPV00000000	*A. aphrophilus*	2008	Brain abscess	456,646	33	51
C2008000870	QEPU00000000	*A. aphrophilus*	2008	Blood	162,253	33	34
C2006002596	QEPT00000000	*H. parainfluenzae*	2006	Blood	246,220	28	47
C2006000788	QEPS00000000	*H. parahaemolyticus*	2006	Bronchial wash	1,108,179	15	72
C2005004058	QEPR00000000	*H. parainfluenzae*	2005	Wound	517,256	16	42
C2004002729	QEPQ00000000	*H. parainfluenzae*	2004	Sputum	466,107	15	53
C2004002727	QEPP00000000	*H. parainfluenzae*	2004	Blood	526,737	23	43
C2004000280	QEPO00000000	*H. parainfluenzae*	2004	Toe	200,727	32	64
C2002001239	QEPN00000000	*H. sputorum*	2002	Throat	470,910	29	36
C2001002503	QEPM00000000	*A. segnis*	2001	Sputum	341,448	13	31
C2001002324	QEPL00000000	*H. haemolyticus*	2001	Sputum	392,601	20	27
C2000002669	QEPK00000000	*A. segnis*	2000	Penile lesion	153,666	22	30

biochemical and phenotypic properties alone due to the shared characteristics among members, while the use of molecular methods for identification has also been problematic due to the high rate of recombination and horizontal gene transfer that occurs between the commensals and pathogens (7). Misidentification of the commensals as the pathogenic species is not uncommon and has been reported at a rate as high as 40% in some clinical labs (8). In recent years, whole-genome sequencing (WGS) has alternatively been used to identify unique genomic targets to discriminate between species in other assays and to provide extensive genomic data that can be used for comparative genomic analysis of *Haemophilus* species (8). The genomic sequences for 7 *Haemophilus* and *Aggregatibacter* species in this study will provide data for future studies examining species delineation and unique genomic targets among *Haemophilus* species.

Data availability. The draft genome sequences have been deposited in GenBank under the accession numbers listed in Table 1.

ACKNOWLEDGMENTS

The Minnesota Department of Health received support by appointment to the Research Participation Program at the Center for Food Safety and Applied Nutrition, U.S. Food & Drug Administration, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Food & Drug Administration. This study was also supported by the CDC Advanced Molecular Detection Initiative (AMD-76).

We declare no competing interests.

REFERENCES

1. Fink DL, Geme JW. 2006. The genus *Haemophilus*, p 1034–1061. *In* Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E (ed), The Prokaryotes, vol6. Proteobacteria: gamma subclass. Springer, New York, NY. https://doi.org/10.1007/0-387-30746-x_40.
2. Musher DM. 1996. *Haemophilus* species. *In* Baron S (ed), Medical microbiology, 4th ed. University of Texas Medical Branch at Galveston, Galveston, TX.
3. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10–12. https://doi.org/10.14806/ej.17.1.200.
4. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.

5. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.
6. Nørskov-Lauritsen N. 2014. Classification, identification, and clinical significance of *Haemophilus* and *Aggregatibacter* species with host specificity for humans. Clin Microbiol Rev 27:214–240. https://doi.org/10.1128/CMR.00103-13.
7. Price EP, Sarovich DS, Nosworthy E, Beissbarth J, Marsh RL, Pickering J, Kirkham L-AS, Keil AD, Chang AB, Smith-Vaughan HC. 2015. *Haemophilus influenzae*: using comparative genomics to accurately identify a highly recombinogenic human pathogen. BMC Genomics 16:641. https://doi.org/10.1186/s12864-015-1857-x.
8. Hu F, Rishishwar L, Sivadas A, Mitchell GJ, Jordan IK, Murphy TF, Gilsdorf JR, Mayer LW, Wang X. 2016. Comparative genomic analysis of *Haemophilus haemolyticus* and nontypeable *Haemophilus influenzae* and a new testing scheme for their discrimination. J Clin Microbiol 54:3010–3017. https://doi.org/10.1128/JCM.01511-16.