Draft Genome Sequences of Three Nontypeable Strains of *Haemophilus influenzae*, C188, R535, and 1200, Isolated from Different Types of Disease

Ulrike Kappler, Rabeb Dhouib, Remya Purushothaman Nair, Alastair G. McEwan

School of Chemistry & Molecular Biosciences, Australian Infectious Diseases Centre, The University of Queensland, St. Lucia, Queensland, Australia

ABSTRACT Nontypeable *Haemophilus influenzae* is a persistent human respiratory pathogen known to be involved in a range of acute and chronic respiratory diseases. Here, we report the genome sequences of three *H. influenzae* strains isolated from sputum, otitis media, and blood. Comparative analyses revealed significant differences in the gene contents including the presence of genes mediating antibiotic resistance.

*H*aemophilus influenzae (HI) is a host adapted human pathogen that usually resides in the respiratory tract where it can either exist as a commensal or cause and contribute to acute or chronic diseases such as otitis media, sinusitis, conjunctivitis, chronic obstructive pulmonary disease (COPD), asthma, and bronchiectasis (1). Recently, an increase in respiratory disease cases caused by nontypeable HI (NTHi) strains has been noted, and this includes more aggressive forms of disease that were previously not thought to be associated with NTHi (1–3).

Despite having small genome sizes (1.7 to 1.9 Mb; ~1,700 to 1,900 genes) which is in keeping with the strong specialization of *H. influenzae* to exclusive growth in the human host, *H. influenzae* strains are known to be genetically variable, with only about 1,400 to 1,500 genes being common to strains (4, 5), and significant numbers of unique genes being found in each strain.

Here, we have sequenced the genomes of NTHi strains isolated from different types of disease: strain C188 is a blood isolate, NTHi 1200 originates from a Finnish study of otitis media, and R535 is a sputum isolate (http://pubmlst.org/hinfluenzae/, accessed 10 January 2017). All three strains have been used in previous studies mostly as parts of strain collections (6, 7), and the availability of complete genome data for these strains will enhance the ability to compare and fully interpret previous and future work (Table 1).

Genomic DNA was isolated using the PureLink genomic DNA kit (Life Technologies, Inc.) and adjusted to 5 ng/mL using a Qubit broad range assay (Thermo Fisher Scientific) before sequencing at the Australian Centre for Ecogenomics (ACE) using the manufacturer's standard protocol for Nextera XT libraries (Illumina). Libraries were pooled at equimolar amounts after quantification with HS D500 Agilent tape (Agilent, TapeStation). Sequencing was performed on NextSeq 500 (Illumina) on a 2 $\times$ 150 PE run with V2 chemistry at a depth of 0.5 Gb per sample. Assembly used Spades v3.9.0 (8) with default parameters for isolate genomes. Read mapping used BamM v1.7.3 (https://github.com/Ecogenomics/BamM) which includes SAMtools (9). Annotation used the NCBI annotation pipeline (NCBI_PGAP) (10).

As expected, the three genomes show the typical variation in the number of protein encoding genes: 1,719 (C188), 1,623 (R535), and 1,743 (1200). A proteinortho (11)

TABLE 1 Properties of the draft genomes for *H. influenzae* strains C188, R535, and 1200

Strain name	Bioproject sample no.	Accession no.	Coverage	No. of contigs	Completeness[a]	Size (Mb)	G+C content (%)	No. of genes	No. of proteins
C188	SAMN05942198	MQMI00000000	455×	56	99.45 (0.07)	1.85	37.9	1,872	1,719
R535	SAMN05938535	MQMJ00000000	545×	61	99.35 (0.23)	1.77	37.9	1,774	1,623
1200	SAMN05942090	MQMH00000000	682×	64	98.86 (0)	1.88	38	1,906	1,743

[a]As determined by CHECKM (13). In parentheses: level of contamination, which is negligible for all three genomes.

comparison of the protein coding genes in *H. influenzae* RD (12) and the three newly sequenced strains revealed that only 1,356 proteins were common to all four strains. Varying numbers of unique proteins (C188: 132; R535: 29; 1200: 136) were present, and these were dominated by hypothetical proteins in all cases (52 to 74%). In addition, unique proteins of phage/transposon origin were particularly abundant in NtHi strain 1200, which also appeared to carry several genes encoding antimicrobial resistance that were notably absent in the other strains, which both originate from an earlier isolation period as well as geographically distinct locations.

Accession number(s). The whole-genome shotgun projects have been deposited in GenBank under the accession numbers specified in Table 1. The versions described in this paper are the first versions.

REFERENCES

1. Van Eldere J, Slack MPE, Ladhani S, Cripps AW. 2014. Non-typeable *Haemophilus* influenzae, an under-recognised pathogen. Lancet Infect Dis 14:1281–1292. https://doi.org/10.1016/S1473-3099(14)70734-0.
2. Langereis JD, de Jonge MI. 2015. Invasive disease caused by nontypeable *Haemophilus influenzae*. Emerg Infect Dis 21:1711–1718. https://doi.org/10.3201/eid2110.150004.
3. Centers for Disease Control and Prevention. 2015. *Haemophilus influenzae*, invasive disease: 2015 case definition. https://wwwn.cdc.gov/nndss/conditions/haemophilus-influenzae-invasive-disease/case-definition/2015/.
4. Hogg JS, Hu FZ, Janto B, Boissy R, Hayes J, Keefe R, Post JC, Ehrlich GD. 2007. Characterization and modeling of the *Haemophilus influenzae* core and supragenomes based on the complete genomic sequences of Rd and 12 clinical nontypeable strains. Genome Biol 8:R103. https://doi.org/10.1186/gb-2007-8-6-r103.
5. Erwin AL, Smith AL. 2007. Nontypeable *Haemophilus influenzae*: understanding virulence and commensal behavior. Trends Microbiol 15:355–362. https://doi.org/10.1016/j.tim.2007.06.004.
6. Erwin AL, Nelson KL, Mhlanga-Mutangadura T, Bonthuis PJ, Geelhood JL, Morlin G, Unrath WCT, Campos J, Crook DW, Farley MM, Henderson FW, Jacobs RF, Mühlemann K, Satola SW, van Alphen L, Golomb M, Smith AL. 2005. Characterization of genetic and phenotypic diversity of invasive nontypeable *Haemophilus influenzae*. Infect Immun 73:5853–5863. https://doi.org/10.1128/IAI.73.9.5853-5863.2005.
7. Kidd SP, Jiang D, Jennings MP, McEwan AG. 2007. Glutathione-dependent alcohol dehydrogenase AdhC is required for defense against nitrosative stress in *Haemophilus influenzae*. Infect Immun 75:4506–4513. https://doi.org/10.1128/IAI.00487-07.
8. 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.
9. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
10. Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity GM, Kodira CD, Kyrpides N, Madupu R, Markowitz V, Tatusova T, Thomson N, White O. 2008. Toward an online repository of standard operating procedures (SOPs) for (meta) genomic annotation. OMICS 12:137–141. https://doi.org/10.1089/omi.2008.0017.
11. Lechner M, Findeiß S, Steiner L, Marz M, Stadler PF, Prohaska SJ. 2011. Proteinortho: detection of (Co-)orthologs in large-scale analysis. BMC Bioinformatics 12:124.
12. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, McKenney K, Sutton G, Fitzhugh W, Fields C, Gocayne JD, Scott J, Shirley R, Liu LI, Glodek A, Kelley JM, Weidman JF, Phillips CA, Spriggs T, Hedblom E, Cotton MD, Utterback TR, Hanna MC, Nguyen DT, Saudek DM, Brandon RC, Fine LD, Fritchman JL, Fuhrmann JL, Geoghagen NSM, Gnehm CL, McDonald LA, Small KV, Fraser CM, Smith HO, Venter JC. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269:496–512. https://doi.org/10.1126/science.7542800.
13. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.