DATA NOTE   **Open Access**

Draft genome sequence of a nontypeable *Haemophilus influenzae* strain used in the study of human respiratory infection

Rajendra KC[1] and Ronan F. O'Toole[2*]

Abstract

Objectives: Nontypeable *Haemophilus influenzae* (NTHi) is an important human respiratory bacterium that can cause a range of diseases including sinusitis, otitis media, conjunctivitis, pneumonia as well as acute exacerbations of chronic obstructive pulmonary disease (COPD). A number of studies have used NTHi clinical isolate RHH-3 as a laboratory strain for experimentation examining the effect of cigarette smoke and more recently, biomass smoke, on the susceptibility and response of cells lining the respiratory tract to infection. Therefore, definition of the genome content of RHH-3 is required to fully elucidate human-NTHi interactions associated with initial infection and subsequent development of respiratory disease.

Data description: Here, we present the draft genome sequence of NTHi RHH-3 collected from the sputum of a patient at the Royal Hobart Hospital, Tasmania, Australia. The assembled genome size was 1,839,376 bp consisting of 61 contigs (> 500 bp), with a G+C content of 38.1%. This draft genome data can be accessed at DDBJ/ENA/GenBank under the accession number JADPRR000000000.

Keywords: Whole genome sequence, Nontypeable *Haemophilus influenzae*, Chronic obstructive pulmonary disease

Objective

Nontypeable *Haemophilus influenzae* (NTHi) strains are common commensal inhabitants of the human nasopharynx. However, they can spread to the sinuses or middle ear via the eustachian tube causing sinusitis and otitis media, respectively, and can also migrate to the eyes causing conjunctivitis [1, 2]. Moreover, they can penetrate into the nasopharyngeal mucosa, or descend to the lower regions of the respiratory tract, resulting in invasive infections that include septicaemia and meningitis, or non-invasive infections such as pneumonia and exacerbations of COPD [3–5]. Environmental factors, such as

exposure to tobacco or biomass smoke have been found to increase susceptibility to infection by respiratory bacteria such as NTHi [6–8]. NTHi strain RHH-3 has been used in mechanistic studies investigating how tobacco and biomass smoke exposure increases the risk of airway infection [9]. The draft assembled genome sequence of NTHi RHH-3 presented here will enable more in-depth studies to be conducted on specific genes that promote NTHi survival and propagation in the COPD lung or that contribute to inflammation that results in tissue impairment and disease. This will provide further insights into the role of NTHi infection in the pathogenesis of COPD. It would also be interesting to investigate in future work whether exposure of lung tissue to smoke predisposes an individual to colonization by a subset of NTHi strains, given a recent finding from pan-genome-wide association analysis that certain NTHi accessory genes are significantly associated with COPD [10].

*Correspondence: r.otoole@latrobe.edu.au
[2] Department of Pharmacy and Biomedical Sciences, School of Molecular Sciences, College of Science, Health and Engineering, La Trobe University, Victoria 3690, Australia
Full list of author information is available at the end of the article

Data description

NTHi strain RHH-3 was isolated from the sputum of a patient presenting with lower respiratory tract infection at the Royal Hobart Hospital, Australia [9, 11]. The sputum sample was homogenized and cultured on chocolate blood agar plates at 35 °C in a CO_2 atmosphere as previously described [12]. Isolated Gram-negative rod colonies, with small and translucent colony morphologies suggestive of *Haemophilus* species, were identified as NTHi through the use of matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS, Bruker Daltonics GmbH, Leipzig, Germany). The isolate was then grown overnight on chocolate agar, incubated at 35 °C with 5% CO_2. A single colony from a chocolate agar plate was suspended in 200 μL PBS and then genomic DNA was extracted using the DNeasy Blood and Tissue Kit (Catalog number 69504; Qiagen, USA). Genomic DNA was further purified using the High Pure PCR Template Preparation Kit (Catalog number 11796828001; Roche, Germany). DNA library preparation was performed using a Nextera XT DNA library preparation kit (Catalog number FC-131-1024; Illumina, USA) as described previously [10, 13, 14]. Sequencing was performed using a MiSeq Reagent Kit v2 (300-cycles) (Catalog number MS-102-2002) with 150-bp paired-end sequencing as previously described [12]. In total, 933,328 paired-end reads were generated, representing an average read depth of 73.13-fold (Table 1). Reads were trimmed of adapters using Trimmomatic [15] and de novo assembly of reads was performed with SPAdes v3.12.0 [16]. All parameters were set to default except for the size of k-mers which were manually chosen as 21, 33, 43, 53, 63, 75. This resulted in the generation of a 1,839,376 bp draft genome consisting of 61 contigs ($\geq$ 500 bp) that covered 82.74% of the *H. influenzae* 86-028NP genome, a well-studied NTHi isolate (Table 2) [17]. The N50 contig was 52,548 bp, and the overall GC content was 38.1% (Table 2). The genome assembly quality, including completeness with respect to the 86-028NP genome, was determined using the QUAST quality assessment tool

Table 2 Genome assembly/annotation statistics

Total sequence length	1,839,376
Number of contigs	138
Contig N50	52,548
Contig L50	12
GC content	38.1%
Coverage of *H. influenzae* 86-028NP genome	82.74%
Total number of genes	1959
Number of protein coding sequences	1907
Number of RNA genes	52

[18]. In addition, the RHH-3 genome was estimated as 99.77% complete with 0% contamination by CheckM [19].

The identity of strain RHH-3 was confirmed by its 16S ribosomal RNA gene sequence (Table 1) (16S rRNA gene sequence, 1543 bp, BLAST identity of 99.48% to *H. influenzae* strain NCTC11931 accession: LS483392.1). The draft sequence of RHH-3 was submitted to the *H. influenzae* multi-locus sequence typing (MLST) website (https://pubmlst.org/hinfluenzae/) for the purposes of generating an in silico MLST profile [20]. The allelic profile of seven housekeeping genes used in the *H. influenzae* MLST was well-defined in RHH-3 *i.e.,* *adk*_98, *atpG*_2, *frdB*_70, *fucK*_15, *mdh*_310, *pgi*_158, and *recA*_4 however, the combination of these alleles was novel as an MLST sequence type corresponding to this allele profile was not available in the *H. influenzae* MLST database (https://pubmlst.org/organisms/haemophilus-influenzae). Based on its unique allele profile, RHH-3 has been assigned an MLST sequence type ST-2380. It is common for NTHi strains to have diverse MLST types due to a relatively high rate of recombination across the genome [21, 22]. Gene prediction and annotation was performed using the Rapid Annotation System Technology (RAST) server [23–25], which identified a total of 1,959 genes consisting of 1,907 protein coding sequences, and 5 rRNA and 47 tRNA genes (Table 2). Default parameters were used for all software unless otherwise specified.

Table 1 Overview of data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Draft genome assembly	NTHi strain RHH-3, whole genome sequence	FASTA	Genbank (https://www.ncbi.nlm.nih.gov/nuccore/JADPRR0000 00000.1/) [26]
16S ribosomal RNA gene, partial sequence	NTHi strain RHH3, 16S rRNA gene sequence	FASTA	Genbank (https://www.ncbi.nlm.nih.gov/nuccore/MW255 938.1/) [27]

Limitations

Comparative analyses were not performed and further investigations are needed to determine the relatedness of RHH-3 to a diverse range of other NTHi isolates.

Declarations

Ethics approval and consent to participate

This work was conducted in accordance with Ethics Approval H0016214 from the Tasmanian Health and Medical Human Research Ethics Committee. No research participants were specifically recruited. An already-collected specimen obtained from routine diagnostic laboratory testing and devoid of patient identifiers was used for this non-interventional retrospective study in which a waiver of consent was applicable.

Consent for publication

Not applicable.

Competing interests

The authors have declared that no competing interest exists.

Author details

[1] Menzies Institute for Medical Research, University of Tasmania, Tasmania 7005, Australia. [2] Department of Pharmacy and Biomedical Sciences, School of Molecular Sciences, College of Science, Health and Engineering, La Trobe University, Victoria 3690, Australia.

References

1. Hu YL, Lee PI, Hsueh PR, Lu CY, Chang LY, Huang LM, Chang TH, Chen JM. Predominant role of *Haemophilus influenzae* in the association of conjunctivitis, acute otitis media and acute bacterial paranasal sinusitis in children. Sci Rep. 2021;11(1):11.
2. Faden H, Bernstein J, Brodsky L, Stanievich J, Krystofik D, Shuff C, Hong JJ, Ogra PL. Otitis media in children. I. The systemic immune response to nontypable Hemophilus influenzae. J Infect Dis. 1989;160(6):999–1004.
3. Murphy TF, Brauer AL, Schiffmacher AT, Sethi S. Persistent colonization by Haemophilus influenzae in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2004;170(3):266–72.
4. Groenewegen KH, Wouters EF. Bacterial infections in patients requiring admission for an acute exacerbation of COPD; a 1-year prospective study. Respir Med. 2003;97(7):770–7.
5. Sunakawa K, Takeuchi Y, Iwata S. Nontypeable *Haemophilus influenzae* (NTHi) epidemiology *Kansenshogaku zasshi*. J Jpn Assoc Infect Dis. 2011;85(3):227–37.
6. Rylance J, Fullerton DG, Scriven J, Aljurayyan AN, Mzinza D, Barrett S, Wright AKA, Wootton DG, Glennie SJ, Baple K, et al. Household air pollution causes dose-dependent inflammation and altered phagocytosis in human macrophages. Am J Respir Cell Mol Biol. 2015;52(5):584–93.
7. van der Vaart H, Postma DS, Timens W, ten Hacken NHT. Acute effects of cigarette smoke on inflammation and oxidative stress: a review. Thorax. 2004;59(8):713–21.
8. Phipps JC, Aronoff DM, Curtis JL, Goel D, O'Brien E, Mancuso P. Cigarette smoke exposure impairs pulmonary bacterial clearance and alveolar macrophage complement-mediated phagocytosis of *Streptococcus pneumoniae*. Infect Immun. 2010;78(3):1214–20.
9. Shukla SD, Fairbairn RL, Gell DA, Latham RD, Sohal SS, Walters EH, O'Toole RF. An antagonist of the platelet-activating factor receptor inhibits adherence of both nontypeable *Haemophilus influenzae* and *Streptococcus pneumoniae* to cultured human bronchial epithelial cells exposed to cigarette smoke. Int J Chron Obstruct Pulmon Dis. 2016;11:1647–55.
10. Kc R, Leong KWC, Harkness NM, Lachowicz J, Gautam SS, Cooley LA, McEwan B, Petrovski S, Karupiah G, O'Toole RF. Whole-genome analyses reveal gene content differences between nontypeable *Haemophilus influenzae* isolates from chronic obstructive pulmonary disease compared to other clinical phenotypes. Microb Genom. 2020. https://doi.org/10.1099/mgen.0.000405.
11. Kc R, Hyland IK, Smith JA, Shukla SD, Hansbro PM, Zosky GR, Karupiah G, O'Toole RF. Cow dung biomass smoke exposure increases adherence of respiratory pathogen nontypeable *Haemophilus influenzae* to human bronchial epithelial cells. Exposure Health. 2020;12:883–95.
12. Kc R, Leong KWC, McEwan B, Lachowicz J, Harkness NM, Petrovski S, Karupiah G, O'Toole RF. Draft genome sequence of an isolate of nontypeable *Haemophilus influenzae* from an acute exacerbation of chronic obstructive pulmonary disease in Tasmania. Microbiol Resour Announc. 2020;9(19):e00375-20.
13. Gautam SS, Rajendra K, Leong KW, Mac Aogáin M, O'Toole RF. A step-by-step beginner's protocol for whole genome sequencing of human bacterial pathogens. J Biol Methods. 2019;6(1):e110.
14. Gautam SS, Mac Aogain M, Cooley LA, Haug G, Fyfe JA, Globan M, O'Toole RF. Molecular epidemiology of tuberculosis in Tasmania and genomic characterisation of its first known multi-drug resistant case. PLoS ONE. 2018;13(2):e0192351.
15. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England). 2014;30(15):2114–20.
16. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455–77.
17. Harrison A, Dyer DW, Gillaspy A, Ray WC, Mungur R, Carson MB, Zhong H, Gipson J, Gipson M, Johnson LS, et al. Genomic sequence of an otitis media isolate of nontypeable Haemophilus influenzae: comparative study with *H. influenzae* serotype d, strain KW20. J Bacteriol. 2005;187(13):4627–36.
18. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29(8):1072–5.
19. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25(7):1043–55.
20. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res. 2018;3:124.
21. Lacross NC, Marrs CF, Patel M, Sandstedt SA, Gilsdorf JR. High genetic diversity of nontypeable *Haemophilus influenzae* isolates from two children attending a day care center. J Clin Microbiol. 2008;46(11):3817–21.

22. Pérez-Losada M, Browne EB, Madsen A, Wirth T, Viscidi RP, Crandall KA. Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data. Infect Genet Evol. 2006;6(2):97–112.

23. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008;9(1):75.

24. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, et al. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep. 2015;5:8365.

25. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, et al. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res. 2014;42(Database issue):D206-214.

26. KC R. O'Toole RF. *Haemophilus influenzae* strain RHH-3, whole genome shotgun sequencing project. GenBank; 2020. The National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/nuccore/JADPRR000000000.1/.

27. KC R. O'Toole RF. *Haemophilus influenzae* strain RHH-3 16S ribosomal RNA gene, partial sequence. GenBank; 2020. The National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/nuccore/MW255938.1/.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.