



The Brazilian Journal of INFECTIOUS DISEASES

www.elsevier.com/locate/bjid



Original Article

Whole-genome analysis of *haemophilus influenzae* invasive strains isolated from Campinas state University hospital. An epidemiological approach 2012 - 2019 and ancestor strains

Rafaella Fabiana Carneiro Pereira, João Paulo de Oliveira Guarnieri^{id},
Carlos Fernando Macedo da Silva^{id}, Bruno Gaia Bernardes,
Marcelo Lancellotti^{id}*

Biotechnology Laboratory, LABIOTEC, Faculty of Pharmaceutical Sciences (FCF), Campinas State University UNICAMP, São Paulo, SP, Brazil

ARTICLE INFO

Article history:

Received 29 August 2021

Accepted 23 November 2021

Available online 24 December 2021

Keywords:

Whole-genome

Haemophilus influenzae

Sequencing

Invasive disease

ABSTRACT

Thirteen *Haemophilus influenzae* invasive strains isolated from patients at Clinical Hospital of State University of Campinas, from May 2013 through August 2019, was submitted to Illumina genome sequencing HiSeq platform. Further *in silico* analysis of serogroup and Multi Locus Sequence Typing (MLST) from whole DNA sequencing had demonstrated the actual clonal distribution in the Campinas Metropolitan region. Thus, results showed the existence of a new ST *Haemophilus influenzae* found in the Brazilian territory and an increase of strains belonging to serogroup a (three strains also belonging to ST23). In conclusion, we observed an increase of non-typable *H. influenzae* (NTHi) and a strain involved in invasive diseases in the Campinas – São Paulo region after frequent detection of those serotypes and genotypes in other Brazilian regions.

© 2021 Sociedade Brasileira de Infectologia. Published by Elsevier España, S.L.U. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Introduction

Haemophilus influenzae is an important pathogen involved in several invasive diseases that might progress to meningitis, septicemia and death. Also, *Haemophilus influenzae* is known as pleomorphic Gram negative coccobacillus classified in six immunological encapsulated strains (a-f) and non-typable *H. influenzae* – NTHi. The *H. influenzae* type b (Hib) is the most invasive type commonly associated with meningitis and

other upper respiratory tract infections in children and adults.^{1,2} The NTHi strains are associated with moderate diseases of the upper respiratory tract and otitis media in children and pneumonia in adults with cystic fibrosis.^{3,4} Along with the introduction of Hib conjugate vaccine, epidemiology of *H. influenzae* has changed in recent years. NTHi and other serotypes of *H. influenzae* has become more prevalent than Hib around the world. Outbreaks at the Clinical Hospital of University of Campinas, São Paulo State – Brazil have shown an increase of invasive strains after vaccination. The use of Illumina platform for bacterial whole-genome construction set up by our group for characterization of Brazilian Pupuric fever *Haemophilus* strains (Pereira *et al.* 2019⁵) was used for

* Corresponding author.

E-mail address: mlancell@unicamp.br (M. Lancellotti).

<https://doi.org/10.1016/j.bjid.2021.101667>

1413-8670/© 2021 Sociedade Brasileira de Infectologia. Published by Elsevier España, S.L.U. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

this purpose. This work aimed to use the Illumina sequencing method for draft-genome to characterize genome structure and *in silico* virulence factors of 13 invasive strains isolated from blood and cerebrospinal fluid. In addition, an approach about the virulence mechanisms and vaccine escape will be explored in this work.

Material and methods

Bacterial strains

Thirteen *H. influenzae* strains were isolated between May 2013 and August 2019 from patients in the Clinical Hospital of State University of Campinas (HC-UNICAMP). All strains were isolated from blood cultures except one Hi2015-6 isolated from cerebrospinal fluid. Hi38 and Hi45 strains had been characterized by Lancellotti *et al.* 2008⁶ and also isolates from the same hospital in 1998. The bacterial strains were grown in chocolate agar plates or BHI supplement with NAD (4 µg/ mL) and hemin (10 µg/ mL) (Kilian, 1976⁷) and incubated at 37°C with 5% CO₂.⁶

Whole-genome sequencing, assembly, and annotation

Genomic DNA was extracted as described and adapted by Cury *et al.* 2014.⁸ The DNA quality analysis and quantification were performed with NanoDrop (NanoDrop® 2000 - Thermo Scientific®). Libraries were prepared with the Nextera XT DNA library preparation kit (Illumina, CA, USA) and sequenced using the Illumina HiSeq 2500 platform (100-bp single-end reads) at the Genomics section of the Life Sciences Core Facility (LaCTAD, Campinas, São Paulo, Brazil). All libraries were multiplexed on one sequencing run. Quality of reads files were evaluated with FastQC v.0.11.7 (Babraham Bioinformatics, Cambridge, UK). Sequencing reads were trimmed, assembled, and annotated in through PATRIC pipeline v.3.5.43 (<https://patricbrc.org/>).⁹ Reads were trimmed by quality (Quality Phred score > 20) and size (> 20 pb) and Illumina adapters sequences were removed using the FastqUtils tool with Trim Galore v. 0.6.1 and Cutadapt v. 2.2. Assembly and annotation

were performed using the tool Genome Comprehensive Analysis with SPAdes v. 3.10.0¹⁰ with default parameters and RAST tool kit (RASTtk),¹¹ respectively.

Capsular operon analysis and MLST determination

Capsular genes were identified using a BLASTn search with Hicap software^{12,13} and Geneious Prime® 2020.1.1 (<https://www.geneious.com>). For the Multi Locus Sequence Typing (MLST) determination, genes sequences for the housekeeping genes *adk*, *atpG*, *adh*, *frdB*, *fucK*, *mdh*, *pgi*, and *recA* were analyzed at *Haemophilus influenzae* MLST website (<https://pubmlst.org/hinfluenzae/>) sited at the University of Oxford¹⁴ for allele and sequence type (ST) assignment.

Virulence genes and antimicrobial resistance genotypes

Virulence factors and acquired resistance genes were assessed using the Virulence Factors Database (VFDB)¹⁵ and ResFinder v. 2.1,¹⁶ respectively.

Results

The sequencing of strains using the HiSeq2500 platform generated a total of 214,376,882 reads for a total of 19 samples. Table 1 summarizes the raw data from the sequencing of 14 samples analyzed. It is observed that the number of reads generated per sample ranged from 16,131,465 to 45,305 (Hi5 and Hi4 samples respectively), with an average percentage above 90% of bases with a phred score = 20. All analyzes performed are attached to this report with all contigs and genomic notations.

Still analyzing the properties and data of the sequenced genomes, the MLST (Table 2) analyses were performed in order to obtain a clonal analysis of the strains isolated in this study. Furthermore, strains isolated between 1997-1998 called ancestor strain (Hi38), strains isolated already in the years 2010, and strains isolated in 2019 (AS's and Hi's) were analyzed with a draft-whole genome. Regarding the MLST analysis of all strains, new alleles were found in our strains

Table 1 – Data of sequencing of *H. influenzae* strains.

Strain	Year	Site	Reads number	% Bases >= Q30	Sequencing average
AS1	2013	Blood	2,034,494	90.78	110.11
AS3	2012	Blood	5,652,459	92.23	305.92
AS4	2012	Blood	531,893	92.78	28.79
AS6	2014	Blood	3,131,435	92.18	169.48
AS11	2014	Blood	6,031,024	92.94	326.41
Hi1	2015	Oropharynx	4,603,214	92.48	249.13
Hi5	2015	Blood	16,131,465	93.6	873.06
Hi6	2015	Cerebrospinal fluid	9,851,427	94.38	533.17
Hi8	2015	Blood	9,353,479	92.92	506.22
Hi9	2015	Blood	15,383,481	94.66	832.57
Hi11	2015	Blood	12,344,279	94.29	668.09
Hi38	1997-1998	Blood	11,381,624	93.7	615.99
Hi45	1997-1998	Blood	10,710,597	93.43	579.67
HiP1	2019	Blood	3,245,950	89.26	175.68
HiX	2019	Blood	10,818,194	93.8	585.50

Table 2 – MLST genes and sequencing type of *H. influenzae* strains.

Strain	Adk	atpG	frdB	fucK	mdh	pgi	recA	ST
AS1	4	15	7	14	78	90	41	634
AS3	1	8	1	14	9	14	13	11
AS4	1	1	1	1	81	21	5	180
AS6	13	16	5	2	3	11	7	23
AS11	42	9	8	2	7	8	4	524
Hi1	14	51	16	48	29	2	31	556
Hi5	13	16	5	2	3	11	7	23
Hi6	13	16	5	2	3	11	7	23
Hi8	3	18	53	2	7	40	10	1813
Hi9	45	1	1	1	1	1	5	1417
Hi11	11	2	15	8	49	26	3	NA*
Hi38	4	17	4	1	2	9	6	4
Hi45	10	14	4	5	4	7	8	6
HiP1	52	11	232*	8	7	1	3	NA
HiX	13	16	5	2	3	11	191*	NA

* New alleles submitted to <https://pubmlst.org/hinfluenzae/>.

isolated in 2019 (*frdB* allele 232 and *recA* allele 191 of HiP1 and HiX, respectively). In addition, new sequences type for *H. influenzae* found in the Hi11, HiP1 and HiX strains (curated on the MLST website) were also determined.

After assembling the genomic drafts, we observed the presence of NTHi and strains belonging to serotype a *H. influenzae* (strains AS6, Hi5 and Hi6) and to the same ST23. In the Fig. 1, the red arrows show the alterations in capsular operon in those strains. The correlation of the lineage considered

elderly - Hi38 was found not to have the same clonal origin (Fig. 1).

Discussion

Genome determination for studying *H. influenzae* strains associated with invasive diseases had been carried out by our group when Pereira *et al.* determined the whole-genome of *Haemophilus influenzae* that caused Brazilian purpuric fever in 2019.⁵ The expertise of genomic bioinformatics platforms made possible the analysis of invasive *H. influenzae* isolated in Clinical Hospital of Campinas State University in this study. This hospital health services covers all the metropolitan area of Campinas with around 3.2 million inhabitants and 20 cities.^{17,18}

Thus, this analysis of bacterial populations is representative the Southwest Brazilian regions and the discovery of new variants of *H. influenzae* identified in this study is an important information for public health considering the identified new ST profile and the presence of serotype a *H. influenzae* in invasive diseases. Data presented in the supplementary material show several virulence genes detected in the strains analyzed in this study. Those virulence factors had been previously tested by our group as reported Pereira *et al.* 2021¹⁹ as expression of genes related with *H. influenzae* biotype aegyptius autotransporters. However, the supplementary analysis about other genes involved with the *Haemophilus* virulence could be a target for next investigations.

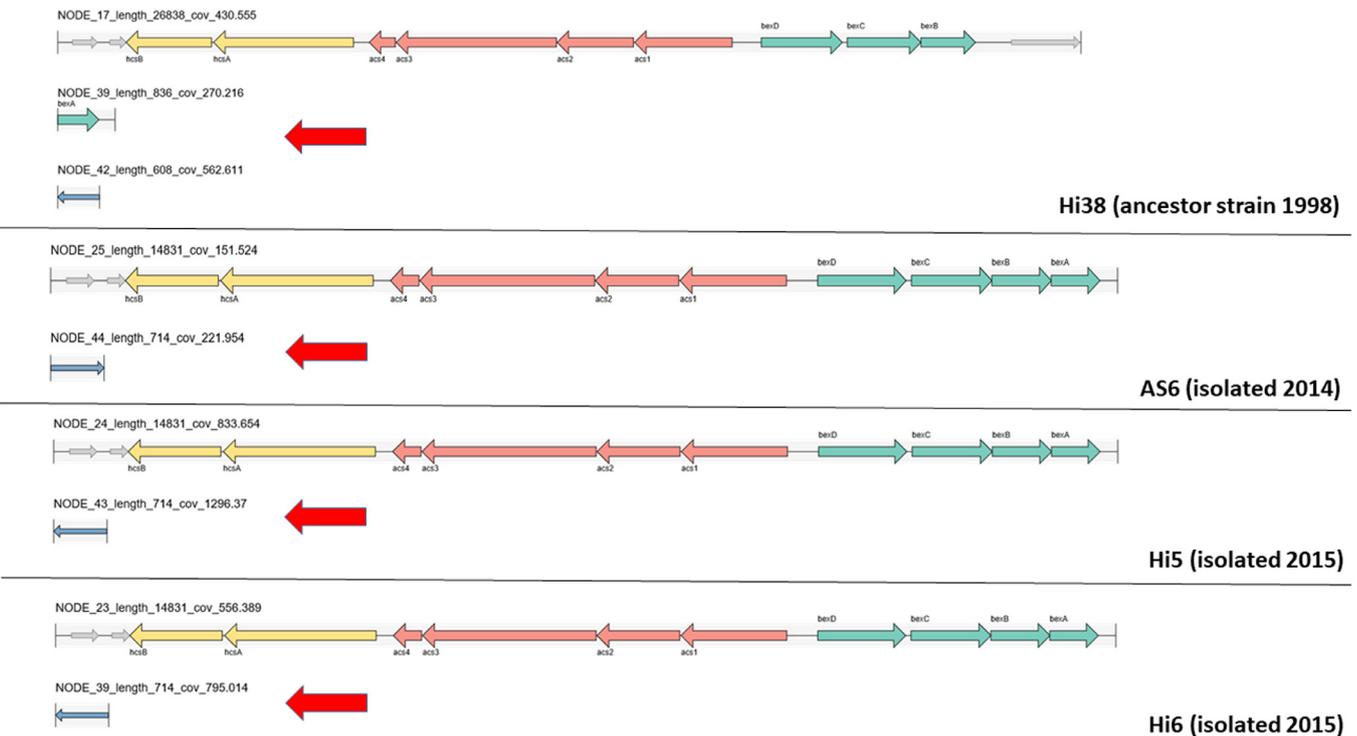


Fig. 1 – Schematic representations of recombination of the capsular operon from *H. influenzae* serotype a comparing the old strain Hi38 (isolated in the 90s) and recent strains AS6, Hi5 and Hi6. The comparison of the strain Hi38 and the recent strains suggest a probable recombination process.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgments

The authors thank the staff of the Life Sciences Core Facility (LaCTAD, UNICAMP) for sequencing the Hae genomes and the biologists of the Clinical Pathology Service, University Hospital (HU), University of Campinas (UNICAMP) for the isolation of the bacterial strains and at technical assistance in all tests. This publication made use of the *Haemophilus influenzae* MLST website (<https://pubmlst.org/hinfluenzae/>) sited at the University of Oxford (Jolley et al. Wellcome Open Res 2018, 3:124 [version 1; referees: 2 approved]). The development of this site has been funded by the Wellcome Trust. This work was supported by grants numbers 2018/09874-7, São Paulo Research Foundation (FAPESP). M.L. are fellowship of grant number 310146/2013-5, CNPq.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.bjid.2021.101667.

REFERENCES

- Kroll JS, Langford PR, Wilks KE, Keil AD. Bacterial [Cu, Zn]-superoxide dismutase: phylogenetically distinct from the eukaryotic enzyme, and not so rare after all!. *Microbiology*. 1995;141:2271–9.
- Giltsdorf JR. *Haemophilus influenzae* non-type b infections in children. *Am J Dis Child*. 1987;141:1063–5. <https://doi.org/10.1001/archpedi.141.10.1063>.
- Hu F, Rishishwar L, Sivadas A, et al. Comparative genomic analysis of *Haemophilus haemolyticus* and nontypeable *Haemophilus influenzae* and a new testing scheme for their discrimination. *J Clin Microbiol*. 2016;54:3010–7. <https://doi.org/10.1128/JCM.01511-16>.
- Hoiby N, Kilian M. *Haemophilus* from the lower respiratory tract of patients with cystic fibrosis. *Scand J Respir Dis*. 1976;57:103–7.
- Pereira RFC, Mofatto LS, Silva ACA, et al. Draft whole-genome sequences of *Haemophilus influenzae* biogroup *aegyptius* strains isolated from five Brazilian Purpuric fever cases and one conjunctivitis case. *Microbiol Resour Announc*. 2019;8. <https://doi.org/10.1128/MRA.00642-19>.
- Lancellotti M, Pace F de, Stehling EG, Villares MCB, Brocchi M, Silveira WD da. Ribotyping, biotyping and capsular typing of *Haemophilus influenzae* strains isolated from patients in Campinas, southeast Brazil. *Braz J Infect Dis*. 2008;12:430–7. <https://doi.org/10.1590/s1413-86702008000500015>.
- Kilian M. A taxonomic study of the genus *Haemophilus*, with the proposal of a new species. *J Gen Microbiol*. 1976;93:9–62. <https://doi.org/10.1099/00221287-93-1-9>.
- Cury GCG, Pereira RFC, de Hollanda LM, Lancellotti M. Inflammatory response of *Haemophilus influenzae* biotype *aegyptius* causing Brazilian Purpuric Fever. *Braz J Microbiol*. 2015;45:1449–54.
- Wattam AR, Davis JJ, Assaf R, et al. Improvements to PATRIC, the all-bacterial Bioinformatics database and analysis resource center. *Nucleic Acids Res*. 2017;45(D1):D535–42. <https://doi.org/10.1093/nar/gkw1017>.
- Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19:455–77. <https://doi.org/10.1089/cmb.2012.0021>.
- Brettin T, Davis JJ, Disz T, 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. <https://doi.org/10.1038/srep08365>.
- Anil A, Spalinskas R, Åkerborg Ö, Sahlén P. HiCapTools: a software suite for probe design and proximity detection for targeted chromosome conformation capture applications. *Bioinformatics*. 2018;34:675–7. <https://doi.org/10.1093/bioinformatics/btx625>.
- Watts SC, Holt KE. hicap: in silico serotyping of the *Haemophilus influenzae* capsule locus. *J Clin Microbiol*. 2019;57.
- 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. <https://doi.org/10.12688/wellcomeopenres.14826.1>.
- Liu B, Zheng D, Jin Q, Chen L, Yang J. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res*. 2019;47(D1):D687–92. <https://doi.org/10.1093/nar/gky1080>.
- Zankari E, Hasman H, Cosentino S, et al. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother*. 2012;67:2640–4. <https://doi.org/10.1093/jac/dks261>.
- Região Metropolitana de Campinas. Wikipédia, a enciclopédia livre. 2021.
- Região Metropolitana de Campinas (RMC) – PDUI RM Campinas. n.d. Available at: https://www.pdui.sp.gov.br/rmc/?page_id=56 [accessed April 4, 2021].
- Pereira RFC, Theizen TH, Machado D, et al. Analysis of potential virulence genes and competence to transformation in *Haemophilus influenzae* biotype *aegyptius* associated with Brazilian Purpuric Fever. *Genet Mol Biol*. n.d.;44. <https://doi.org/10.1590/1678-4685-GMB-2020-0029>.