

RESEARCH

Open Access



Whole-genome sequencing of non-typeable *Haemophilus influenzae* isolated from a tertiary care hospital in Surabaya, Indonesia

Made Ananda Krisna^{1,2*}, Lindawati Alimsardjono³, Korrie Salsabila^{1,6}, Naritha Vermasari³,
Wa Ode Dwi Daningrat^{1,4}, Kuntaman Kuntaman³, Odile Barbara Harrison⁵,
Martin Christopher James Maiden² and Dodi Safari^{1*}

Abstract

Background *Haemophilus influenzae* causes life-threatening invasive diseases such as septicaemia and meningitis. Reports on circulating *H. influenzae* causing invasive disease in lower-middle income settings, including Indonesia, are lacking. This study describes the serotype distributions and whole-genome sequence (WGS) data of *H. influenzae* isolated from hospitalized patients at Soetomo Hospital, Surabaya, Indonesia.

Methods *H. influenzae* isolates were isolated from blood and pleural fluid specimens and identified using culture-based and molecular methods, followed by serotyping and WGS using RT-PCR and Illumina MiSeq, respectively. Sequencing reads were assembled, and further analyses were undertaken to determine the genomic content and reconstruct the phylogeny. A second dataset consisting of publicly available *H. influenzae* genomes was curated to conduct phylogenetic analyses of isolates in this study in the context of globally circulating isolates.

Results Ten *H. influenzae* isolates from hospitalized patients were collected, and septicaemia was the most common diagnosis ($n=8$). RT-PCR and WGS were performed to determine whether all the isolates were nontypeable *H. influenzae* (NTHi). There were four newly identified STs distributed across the two main clusters. A total of 91 out of 126 virulence factor (VF)-related genes in *Haemophilus* sp. were detected in at least one isolate. Further evaluation incorporating a global collection of *H. influenzae* genomes confirmed the diverse population structure of NTHi in this study.

Conclusion This study showed that all *H. influenzae* recovered from invasive disease patients were nonvaccine-preventable NTHi isolates. The inclusion of WGS revealed four novel STs and the possession of key VF-associated genes.

Keywords *Haemophilus influenzae*, Invasive disease, Whole-genome sequencing, Population genetics, Indonesia

*Correspondence:

Made Ananda Krisna
made.krisna@biology.ox.ac.uk
Dodi Safari
dodi004@brin.go.id

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Introduction

Haemophilus influenzae, a pleomorphic gram-negative coccobacillus, frequently resides within the upper respiratory tract and is a common etiological agent of invasive and noninvasive bacterial infections [20]. *Haemophilus influenzae* type b (Hib) was the most prevalent cause of bacterial meningitis in children aged <5 years prior to the introduction of the Hib-conjugate vaccine [36]. In the postvaccination era, the prevalence of Hib disease has greatly decreased [36], however, nontypeable *H. influenzae* (NTHi) has emerged as the predominant cause of *H. influenzae* invasive disease [17, 22]. This pattern is observed across different regions worldwide. Reports from the European Centre for Disease Prevention and Control (ECDC) showed that most *H. influenzae* invasive disease cases were attributed to NTHi (78%) or non-Hib encapsulated isolates (13%) [21, 45]. According to a surveillance program in South Africa in 2018, 64% of invasive *H. influenzae* infections were due to NTHi [27]. Additionally, data from the United States indicated that NTHi has the highest incidence and case fatality rates (CFRs) in the US compared to Hib and non-Hib encapsulated serotypes [39]. Globally, invasive NTHi infection had a greater CFR (17–21%) than invasive Hib infection [37].

The Hib vaccine was introduced in the routine childhood immunization schedule in Indonesia in 2013 [31]. At the time of writing (November 2023), there was a relative lack of data on invasive *H. influenzae* infection post-Hib vaccination from the WHO South East Asia Region (SEARO), including Indonesia [37]. This scarcity of data was partly due to the limited capacity of microbiology laboratories to identify and isolate fastidious organisms. As part of a US Centers for Disease Control and Prevention (CDC)-funded capacity-building project aiming to improve the detection of fastidious organisms, including *H. influenzae*, this study is the first to describe the serotype distribution and population genomics of *H. influenzae* isolated from hospitalized patients in Surabaya, Indonesia, in 2019.

Methods

Haemophilus influenzae isolation and identification

This study used archived isolates and data from a previous project, in which routine sterile-site specimens were collected at Dr. Soetomo Academic General Hospital, a tertiary referral hospital in Surabaya, Indonesia, from January to December 2019. The specimens were inoculated into blood culture bottles and incubated in a BACTEC Incubator (BD BACTEC FX) per the manufacturer's protocol. Positive culture bottles were streaked onto supplemented chocolate agar and incubated at 37°C with 5% CO₂. Suspected *H. influenzae* colonies were

characterized as having a non-haemolytic, opaque to gray color, creamy texture with a pungent indole smell. All suspected colonies were stored in STGG media and frozen at -80°C. The isolates were shipped to the Eijkman Institute for Molecular Biology (Jakarta, Indonesia) on dry ice for further identification by Gram staining and X&V-dependent tests [32]. Real-time PCR (RT-PCR) for *hpd* gene detection was employed to distinguish *H. influenzae* from *H. hemolyticus* and for serotyping (Additional File 1: Supplementary Table 1) [43]. The confirmed isolates were subjected to serotyping by RT-PCR [28, 46].

Whole-genome sequencing and analyses

Whole-genome sequencing (WGS)

Bacterial DNA was extracted using the DNeasy Blood and Tissue® Kit according to the manufacturer's protocol (Qiagen, Carlsbad, CA, USA). The Nextera XT DNA Library Prep® kit was used for library preparation before paired-end WGS (2x250) on a MiSeq (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. FastQ reads were evaluated for quality and trimmed using FastQC v.0.11.9 and Trimmomatic v.0.32 pipeline [2, 6]. Draft genomes were assembled utilizing the SPAdes v.3.15.5 toolkit with the setting optimized for 2x250 bp reads [30]. Quality assurance and quality control (QA/QC) for draft genome assemblies were verified using QUAST. Statistics for assembly quality (number of contigs, genome length, GC content, N50, and L50) are reported in Additional File 1: Supplementary Table 2 [16].

Genome annotation and characterization

Draft genome annotation was performed with Prokka v1.14.6 using default settings. To evaluate the assembly and annotation quality further, one representative of the clinical isolates (Genome ID 1) was aligned to the NCBI *H. influenzae* reference genome (NZ_CP007470.1) using Mauve. The draft genome contigs were reordered based on the alignment and visualized using Genovi and Proksee to mark the positions of important loci (i.e., MLST and rMLST loci) and GC skews in comparison to the reference genome.

Two additional specific annotation steps were employed for the capsule region and genes encoding virulence factors (VFs). The former was performed with the Hicap suite, an *in silico* algorithm for predicting capsule type [44]. The annotation of VF-related genes was performed based on BLAST results from the Virulence Factor Database (VFDB) [24, 29]. Virulence factors are defined as “gene products that enable a microorganism to: 1) colonize a host niche, 2) proliferate, and, 3) cause tissue damage or systemic inflammation” [24], [8]. In the VFDB, VF-related genes for *Haemophilus* sp. were

curated from several reference genomes: *H. influenzae* 86-028NP (NC_007146), PittEE (NC_009566), PittGG (NC_009567), Rd KW20 (NC_000907), *H. ducreyi* 35000HP (NC_002940), *H. somnus* 129T (NC_008309) and 2336 (NC_010519). The first three *H. influenzae* reference genomes were NTHi.

Sequence typing profile assignment

Genome assemblies and relevant, deidentified isolate characterization data were uploaded to the PubMLST website (<http://pubmlst.org>) [19]. All the genomes were evaluated for definitive species identification and potential contamination by characterizing the ribosomal multilocus sequence typing (rMLST) profile. rMLST is a typing method that unifies bacterial genealogy and typing by cataloging variations in 53 bacterial ribosome protein subunit (*rps*) genes. This approach is capable of categorizing bacterial sequences into taxonomic groups at all hierarchical levels [18]. The seven-locus MLST profile and clonal complex (CC) were automatically assigned on the PubMLST website.

Phylogenetic analysis

The genomic diversity of the isolates in the study was evaluated with phylogenetic analyses. PIRATE was used to generate a core-genome alignment from annotated draft genomes [5, 34]. A maximum likelihood (ML) tree was constructed utilizing RaXML v8 and ClonalFrameML to account for recombination events [13, 40]. The resulting tree was visualized and annotated with Microreact [4]. Analyses were implemented on two datasets. The first dataset consisted of ten *H. influenzae* genomes isolated from clinical samples. The second dataset included 506 high-quality, publicly available NTHi genomes from the PubMLST database (<https://pubmlst.org/organisms/haemophilus-influenzae>) in addition to the ten genomes

from this study to evaluate the diversity in the context of previously published genomes (Additional File 1: Supplementary Figure 1 and Additional File 2).

Results

All 10 *H. influenzae* isolates causing invasive diseases at Dr. Soetomo Hospital were nontypeable

Ten *H. influenzae* isolates from ten patients were identified from sterile site specimens (Table 1). All the isolates were positive for the *hpd* gene and negative for capsule-encoding genes. These RT-PCR-based identification and serotyping results were fully concordant with the WGS-based analyses. *H. influenzae* species identification was confirmed based on the rMLST score without any evidence of contamination (Additional File 1: Supplementary Table 3). This score represented the percentage of *rps* gene variation detected in a particular genome supported by the preexisting variation from genomes with a known species [18]. Prediction with the Hicap suite also confirmed that all the isolates were NTHi [44], [11].

All ten *H. influenzae* draft genomes passed the QA/QC step (Additional File 1: Supplementary Table 2). As a further evaluation of the assembly and annotation quality, one of the ten *H. influenzae* genomes (Genome ID 1) was visualized schematically as a circular genome (Fig. 1). The pattern of GC skewness and position of the 7 MLST and 53 rMLST loci visualized from the draft genome highly resembled those of the reference genome (Additional File 2: Supplementary Figure 2 and 3).

Each of the ten isolates possessed different STs with four newly identified STs (STs 2571, 2579, 2576, and 2563). STs 2576 and 2563 were not part of any CCs, while STs 2571 and 2579 belonged to the CC 746 and CC 3, respectively. Among the six isolates with previously known STs, four were part of the CC 107 (STs 485, 503,

Table 1 Demographic and basic clinical characteristics of patients with invasive *H. influenzae* infection at Soetomo Hospital from January – December 2019

No (Genome ID)	Age group	Sex	Specimen type	Month of specimen collection	Place of specimen collection
1	geriatric	female	blood	March	In-patient wards at Soetomo Hospital
2	geriatric	male	blood	April	
3	geriatric	female	blood	April	
4	paediatric	female	blood	May	
5	adult	male	pleural fluid	June	
6	adult	female	blood	July	
7	geriatric	male	pleural fluid (empyema)	August	
8	adult	male	bacteraemia	January	
9	paediatric	female	blood	December	
10	paediatric	female	blood	January	

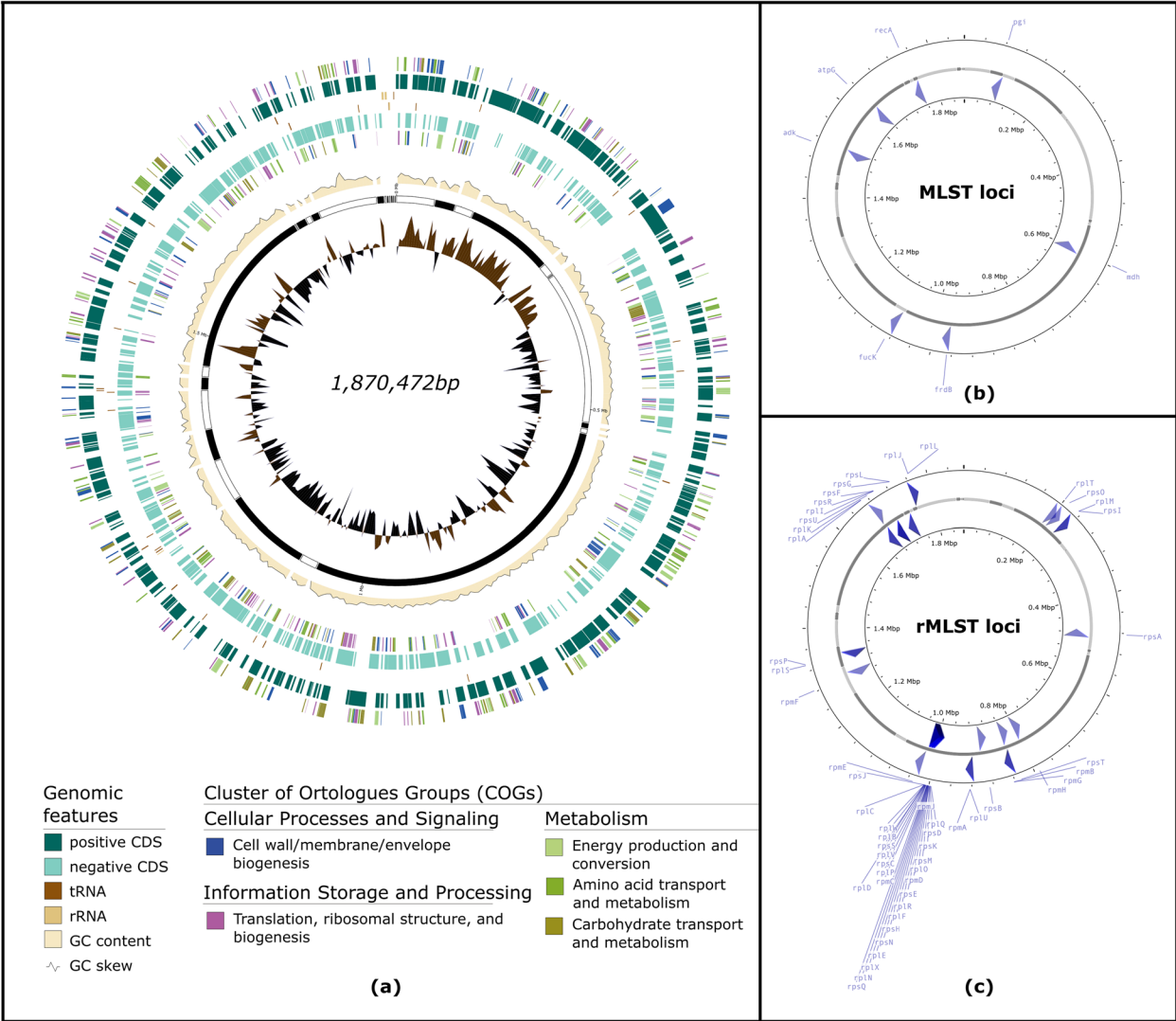


Fig. 1 Circular schematic representation of draft *H. influenzae* genome isolated from invasive clinical isolates (Genome ID 1). The draft assembly was aligned to the complete reference genome *H. influenzae* 477 (NZ_CP007470.1) and its contigs were re-ordered based on the alignment using Mauve. The annotated draft genome was visualized as a circular genome using Genovi and Proksee, genome visualizers for bacteria and archaea. All CDSs were annotated and those belonging to the five most frequent clusters of orthologous groups (COG) were indicated (a). The position of the 7 (b) and 53 loci (c) in the MLST and rMLST scheme, respectively, were also shown

819, and 2036) with the remaining two belonged to CC 396 (ST 695) and CC 11 (ST 103).

Two distinct clusters were observed among the 10 *H. influenzae* genomes

According to the pangenome analysis of ten annotated *H. influenzae* genomes, there were 2,384 genes in total, 1,424 of which were shared. Subsequently, the core genome alignment from the analysis was utilized to construct an ML phylogenetic tree, which inferred their genetic relatedness (Fig. 2).

Two main clusters, C1 and C2, were observed following phylogenetic analyses. All four isolates in C1 belonged to the same CC, the ST-107 complex, with short branch lengths. For C2, no members in this cluster belonged to the same CC, and each had a long branch from the most recent common ancestor. Additionally, all four isolates with newly identified STs were part of C2.

Out of the four isolates clustered in C1, two originated from geriatric (≥ 60 years old) patients, one from an adult (18-60 years old), and one from a paediatric patient (< 18 years old). Only one of these patients was male. There was a total of four isolates from adult patients, three of

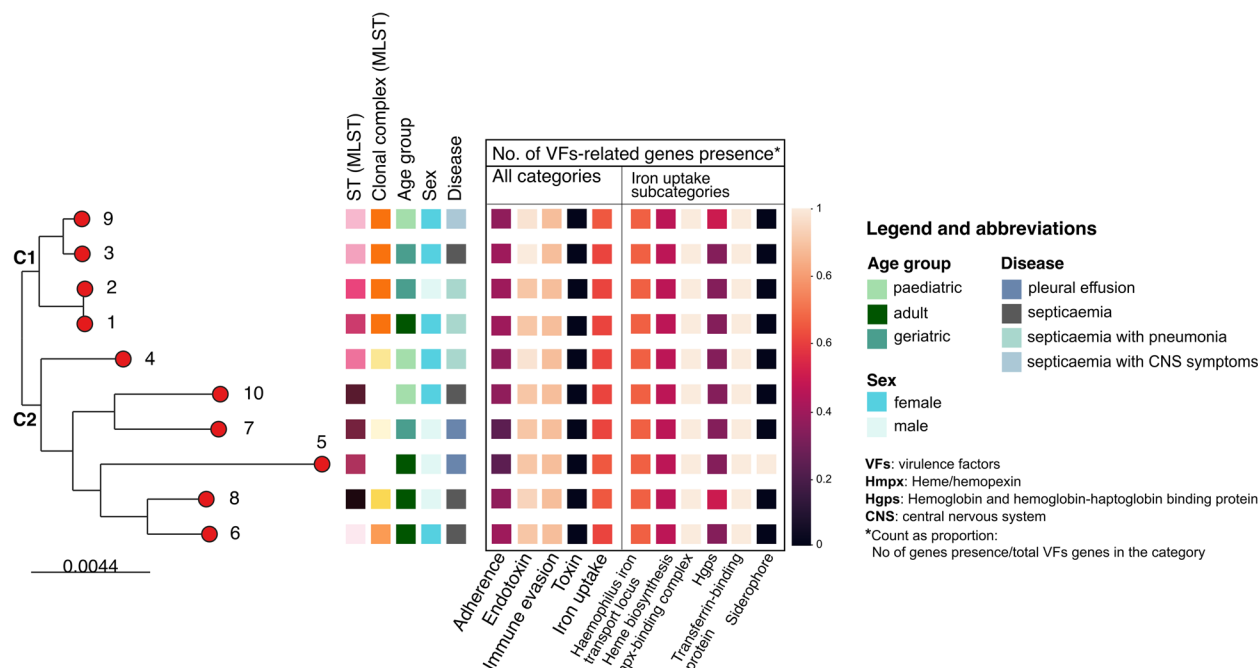


Fig. 2 Maximum-likelihood phylogenetic tree of core-genome alignment of ten *H. influenzae* isolated from hospitalized patients in Surabaya, Indonesia from January to December 2019. Numbers alongside the nodes were genome IDs, as shown in Table 1. The tree was annotated with metadata blocks (from left to right): sequence-type, clonal group, age group, sex, and disease. The metadata blocks in the black square were the proportion of VFs-related genes present in each genome for each VF category (i.e. 1 meaning all VF-s genes in a category were present in a particular genome). The same depiction was generated for the “iron uptake” subcategories

which were grouped in C2 [35]. In contrast to C1, C2 had an equal number of isolates from male and female patients ($n = 3$ each). Although the total number of isolates was small, all patients with pleural effusion ($n = 2$) belonged to C2. Additionally, two out of three patients with both septicaemia and pneumonia were closely clustered within one of the C1 branches (Fig. 2).

The majority of well-defined virulence factor-related genes were present in the ten *H. influenzae* genomes

Among the 126 genes associated with known or predicted VFs in the *Haemophilus* sp. reference genome, 91 (72%) were present in at least one of the Indonesian genomes (Additional File 3). Among the 91 VF-related genes, more than half ($N=51$) were involved in lipooligosaccharide (LOS) production; only three LOS-associated genes, which encode glycosyltransferase family 52 protein (*orfO*), phosphomannomutase (*yhxB/manB*), and hypothetical protein (*lex2A*), present in all the NTHi reference genomes were absent in our collection of genomes (Fig. 3). A total of 73/91 (80.2%) VF-related genes were shared by all the genomes in the present study.

The presence of VF-related genes was also assessed in the context of the phylogenetic analyses (Fig. 2).

There were a consistent number of genes present for each VF category in both clusters; however, there were two C2 isolates (Genome IDs 5 and 7) with a lower number of total adherence-associated genes. These isolates did not possess any *hmwA* or *hmwB* genes (Additional File 1: Supplementary Figure 4). The two isolates were detected in two patients with pleural effusion but without septicaemia. On the other hand, the third isolate obtained from a patient with septicaemia (Genome ID 10) without the *hmwA/B* gene had four additional adherence-related genes, *hifA/B/C/D*, which were not present in the rest of the isolates in the dataset (Additional File 1: Supplementary Figure 4).

In addition to the 126 VF-related genes commonly identified in *Haemophilus* sp., a virulence gene originating from *Acinetobacter baumannii*, *basG*, was detected in one genome (Genome ID 5). This gene encodes histidine decarboxylase, an enzyme responsible for acinetobactin biosynthesis, which functions as a siderophore [26]. There was no prior evidence of the gene being present in several *Haemophilus* sp. reference genomes. Therefore, the isolate was likely to have acquired this gene via horizontal gene transfer (HGT).

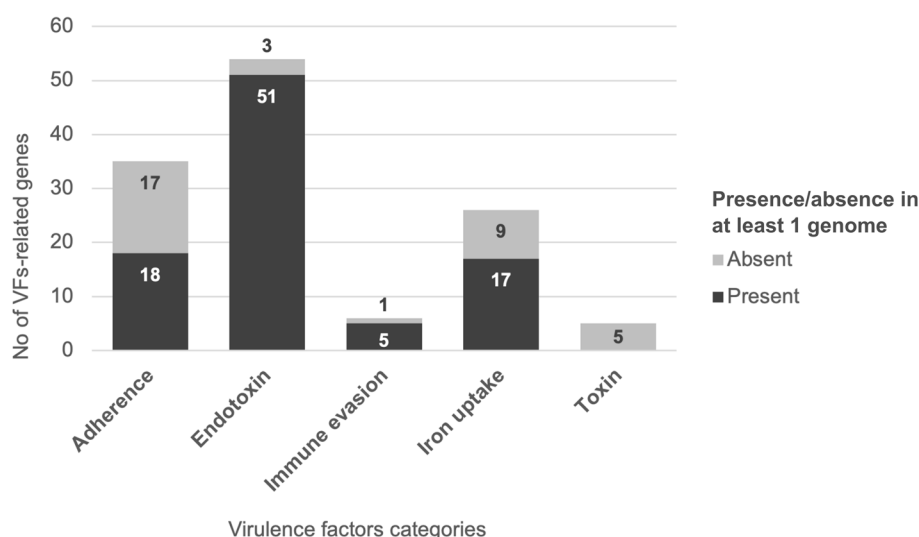


Fig. 3 The number of virulence-associated genes present in at least one *H. influenzae* isolated from hospitalized patients in Surabaya, Indonesia from January to December 2019, grouped by virulence factors (VFs) categories

Globally circulating *H. influenzae* show a diverse population structure, with each *H. influenzae* genome from Soetomo Hospital located across the phylogeny

A total of 506 publicly available NTHi genomes were chosen to evaluate the 10 genomes from this study in the context of the global NTHi population structure. The workflow for choosing these public genomes is depicted in Additional File 1: Supplementary Figure 1. Most (375/506, 74%) genomes originated from Europe, 108 (21%) from North America, and the rest from Oceania (11, 2.2%), Africa (8, 1.5%), and Asia (4, 0.8%). A total of 405 genomes (80%) had associated provenance data of the specimen type from which they were isolated, with the most prevalent being blood (28%), sputum (27%), nasal/nasopharynx swabs (13%), and cerebrospinal fluid (11%). (Additional File 2)

Phylogenetic analysis generated from the core genome nucleotide alignment was performed for a total of 516 genomes. There was no phylogeographic structure with the Indonesian isolates sequenced in this study dispersed across the tree (Fig. 4). Two findings were noteworthy. First, five STs (STs 140, 259, 398, 423, and 914) in the ST-3 and ST-139 CCs putatively resembled ancestral genotypes, consistent with the presence of short branches on the mid-rooted tree. Additionally, a higher degree of genomic diversity was present in isolates belonging to these CCs. Second, genomes with STs not clustering with either ST-3 or ST-139 CCs showed a more clonal structure, indicative of genotypes that diverged from the ancestral state. Indonesian isolates from this study were found either close to the putative ancestral lineage or in clonal groups at the tips of the branches.

Discussion

Although confined to one referral hospital in the country, this study is the first surveillance report of *H. influenzae* invasive disease in Indonesia. Building the capacity to isolate and conduct microbiological identification of the bacterium was essential for this project. Once capacity was in place, WGS was incorporated into the analysis to provide more insights into the population structure of circulating bacteria causing invasive disease. This approach strengthened the microbiological capacity and ensured that WGS analyses could be incorporated with high confidence due to the presence of robust sample processing.

All (10/10) *H. influenzae* isolated from sterile clinical specimens in this study were NTHi. Consistent with multiple epidemiology reports worldwide, after more than two decades of Hib vaccine implementation, NTHi is the major cause of invasive *H. influenzae* disease. The central public health laboratory in the province of Ontario, Canada, and the national referral laboratories in Paraguay and Argentina reported NTHi as the underlying type for 62.3%, 57.9%, and 44.5% of invasive *H. influenzae* disease, respectively [1, 14, 23]. However, there are unique findings in particular regions, such as *H. influenzae* type a as the dominant etiology for invasive diseases (50%) in Northwestern Ontario, Canada and the re-emergence of Hib in Argentina (41.1%) [14], [15]. These reports mostly come from higher-income settings and/or areas outside the WHO South East Asia Region (SEARO), as at the time of the writing, there were limited data on *H. influenzae* infection in the lower-middle income setting, especially in SEARO. Nevertheless, we cannot comment

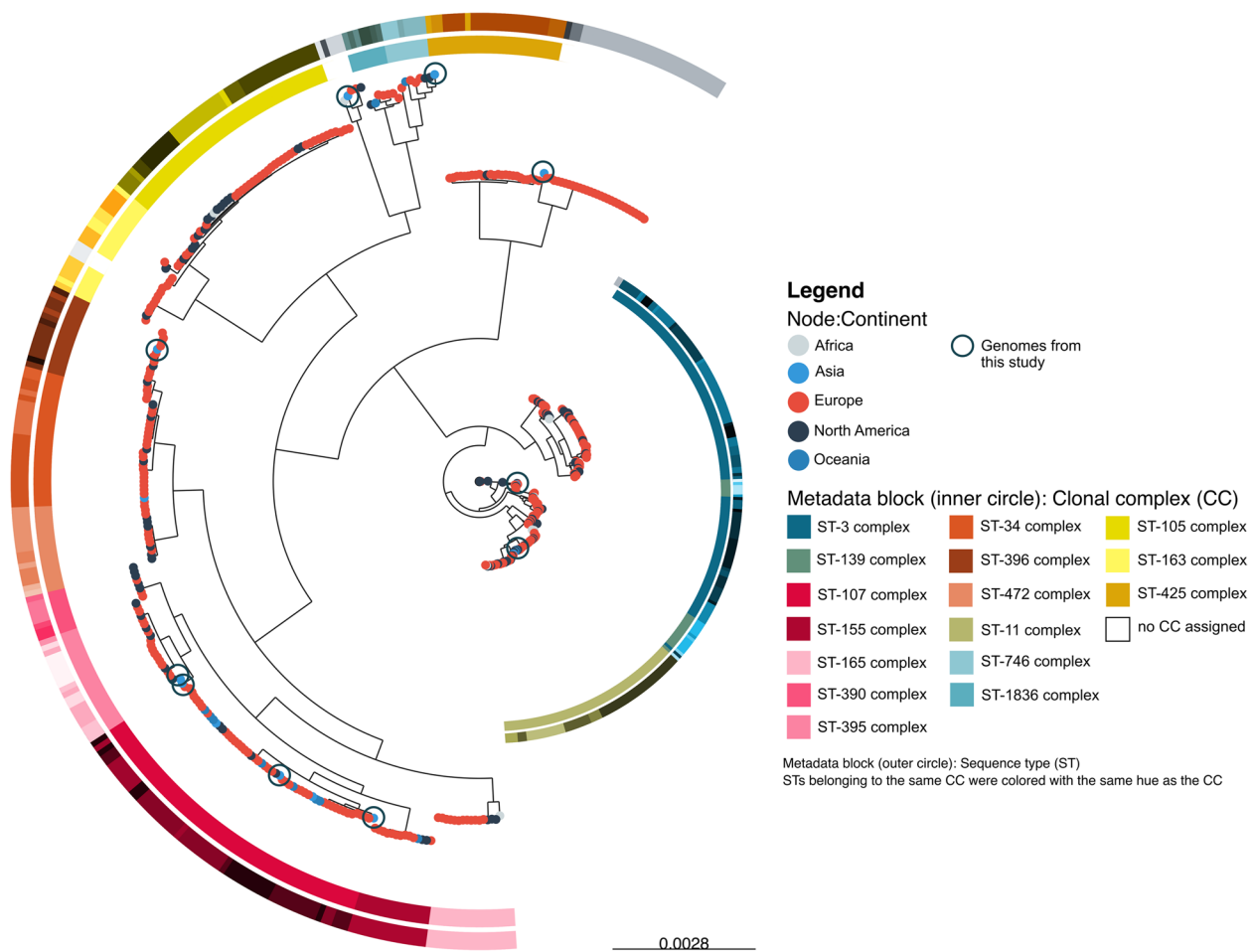


Fig. 4 The population structure of 10 invasive NTHi isolates from Surabaya, Indonesia reflects the population structure of the selected global NTHi genomes collection. The ML tree was constructed based on the core genome alignment and had been corrected for recombination events. The nodes were coloured based on continents, while the inner and outer metadata blocks were CC and ST, respectively. Encircled nodes were the genomes from this study

on the invasive *H. influenzae* disease trend in the post-Hib vaccination era in Indonesia, as this is the first-ever report of this disease in this country.

Most (7/10) patients with *H. influenzae* disease were adults, including four elderly patients (defined as individuals ≥ 60 years old, per the Indonesia Presidential Legislation for National Strategy for Older People) [3]. This finding is in accordance with the known age distribution of invasive NTHi disease in the post-Hib vaccination era. Surveillance reports from the ECDC in 2007–2014 revealed similar findings, with a median age of 58 years for patients with invasive NTHi infections [45]. In Spain, over a study period dating from 2008 to 2019 at a tertiary care center for adults, 65.8% of invasive NTHi patients were > 65 years old [7]. Data from the Queensland Public Health Microbiology Laboratory from 2001 to 2015 demonstrated similar findings, with a median age of 50.5 years [41].

Two patients with invasive NTHi disease in the present study were children under five years old, one of whom was a newborn. A South African surveillance program revealed that the incidence of invasive NTHi was highest in infants [37], and a report from Queensland revealed that 36.5% (54/148) of the cases were in this age group [10]. Additionally, 42% of under five invasive *H. influenzae* infections in Japan were detected in infants [42]. Regardless of geographical region, non-typeable *H. influenzae* has been mentioned as an emerging cause of neonatal sepsis, which was the clinical diagnosis of invasive NTHi disease in newborns in this study.

The diversity observed among the ten NTHi isolates in this study was consistent with the established knowledge of the population structure of this bacterium. Among the 5,345 isolate records in PubMLST database (accessed 13th November 2023), there were fewer than 14 isolates in the database belonging to 9 of the 10 STs

identified in this study. Although an underlying sampling bias cannot be excluded, the *H. influenzae* isolates recovered in this study may reflect the emergence of NTHi genotypes belonging to the ST-107 clonal complex.

The pattern that emerged in the phylogenetic analyses has been reported previously in studies evaluating the population structure of invasive NTHi isolates. Phylogenetic analysis based on core genome single nucleotide polymorphisms (SNPs), or core SNPs, of 58 invasive NTHi isolates from a single tertiary care center in Spain revealed that 30 of them shared one most recent common ancestor with short branches, while the rest displayed high genetic diversity [7]. A report from Denmark's National Surveillance Laboratory of 503 isolates also revealed a similar population structure [38]. This observation suggested that there are two distinct groups of invasive NTHi isolates based on their degree of genome diversity. However, genetic relatedness investigations of invasive clinical *H. influenzae* isolates from Queensland, Australia, did not reveal this distinction [41]. These inconsistencies may be partly caused by the use of core-SNP-based phylogenetic analyses, which rely on the use of an appropriate reference genome, rather than employing *de novo* assembled core genome alignments [33]. Nevertheless, *H. influenzae*, especially NTHi, has been described for its capacity to undergo horizontal gene transfer and recombination events, contributing significantly to diversification [25].

To date, there has been only one in-depth, complete analysis of *H. influenzae* virulence gene profiles, which included 88 invasive NTHi isolates from Portugal [29]. Similar to our findings, this study revealed that the presence of one adhesin-associated gene, *hmwA*, was almost always accompanied by *hmwB*. A similar pattern was observed for invasive NTHi isolates from Spain [7]. The *hmwA* and *hmwB* genes encode two proteins in the two-partner secretion (TPS) pathway in the type V secretion system (TVSS). The concurrent presence of *hmwA* and *hmwB* genes observed in the current and previous studies might indicate the interconnected function of their products [9]. In addition, a study reported a possible immune modulation role of these products in *H. influenzae* [12]. Although there was no further evaluation of how this characteristic relates to clinical manifestations of infection, the resulting compromised immune system could result in a more widespread infection, such as septicemia. The descriptive nature of both current and previous studies prevents any conclusion from being drawn on the relationship between virulence gene presence and clinical outcome. Further *in silico* and *in vitro* research with a more comprehensive dataset is needed to unravel the complex

interplay between the virulence gene repertoire and pathogen population dynamics and its potential impact on clinical manifestations.

Conclusion

This study characterized 10 confirmed NTHi isolates from 10 invasive disease patients in a tertiary referral hospital in Surabaya, Indonesia. The isolation and identification of fastidious bacteria were made possible through a capacity-building project and can feasibly become a part of a routine surveillance program. Integrated WGS analyses revealed four novel STs and added value to understanding the population genomics and biology of the bacterium.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12879-024-09826-8>.

Additional file 1. Supplementary tables and figures for the article "Whole-genome sequencing of Nontypeable *Haemophilus influenzae* isolated from a tertiary care hospital in Surabaya, Indonesia".

Additional file 2. Global dataset metadata.

Additional file 3. Virulence factors presence-absence of 10 non-typeable *Haemophilus influenzae*.

Acknowledgments

The archived specimen in this study was obtained from the capacity-building project funded by the Grant or Cooperative Agreement Number NU2GGH001852-03, funded by the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention or the Department of Health and Human Services. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. We would like to thank Fabiana C. Pimenta, Maria da Gloria Carvalho, Tamara Pilishvili, and Jennifer R. Verani from the National Center for Immunizations and Respiratory Diseases CDC for their expertise and assistance in the capacity-building project essential to this study and feedback for the manuscript, respectively. We would also like to thank all laboratory staff from the Microbiology Laboratory, Dr. Soetomo, Hospital, Wisnu Tatroji, Wisiva Tofriska Paramaiswari, Miftahuddin Majid Khoeri, and Sarah Fitria for their involvement in the capacity-building project.

Authors' contributions

D.S. designed the study and wrote the main manuscript text with M.A.K., O.B.H., and M.C.J.M.; L.A., K.S., N.V., K.K. collected the isolates from clinical samples and conducted microbiological identification; W.O.D.D. coordinated the capacity building activities and collected associated epidemiological data with L.A. and N.V.; M.A.K. and K.S. did confirmatory testing, serotyping and WGS; M.A.K. performed the genomic analyses and created all figures with O.B.H. and M.C.J.M.; all authors reviewed the manuscript.

Funding

The archived specimen in this study was obtained from the capacity-building project funded by the Grant or Cooperative Agreement Number NU2GGH001852-03, funded by the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention or the Department of Health and Human Services. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article and its additional files. The sequence reads were deposited in the NCBI Sequence Read Archive under the project accession number PRJNA1043690.

Declarations

Ethics approval and consent to participate

This study used archived bacterial isolates from a previous project and was approved by the ethical committee of the Eijkman Institute Research Ethics Commission No. 130. Informed consent was not needed prior to data collection since this study was conducted on routinely collected, archived specimens and therefore was waived.

Consent for publication

Not applicable

Competing interests

The authors declare no competing interests.

Author details

¹Eijkman Research Centre for Molecular Biology, National Research and Innovation Agency, Cibinong, West Java, Indonesia. ²Department of Biology, University of Oxford, Oxford, UK. ³Department of Clinical Microbiology, Dr. Soetomo Academic General Hospital, Surabaya, Indonesia. ⁴Centre for Genomic Pathogen Surveillance, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK. ⁵Nuffield Department of Population Health, University of Oxford, Oxford, UK. ⁶Graduate School of Medical and Pharmaceutical Sciences, Chiba University, Chiba, Japan.

Received: 12 March 2024 Accepted: 28 August 2024

Published online: 03 October 2024

References

- Adam HJ, Richardson SE, Jamieson FB, Rawte P, Low DE, Fisman DN. Changing epidemiology of invasive *Haemophilus influenzae* in Ontario, Canada: Evidence for herd effects and strain replacement due to Hib vaccination. *Vaccine*. 2010;28(24):4073–8. <https://doi.org/10.1016/j.vaccine.2010.03.075>.
- Andrews, Simon (2010) *FastQC: A Quality Control Tool for High Throughput Sequence Data*. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Anon (2021) *Peraturan Presiden (PERPRES) Nomor 88 Tahun 2021 tentang Strategi Nasional Kelanjutan*. <https://peraturan.bpk.go.id/Home/Details/178090/perpres-no-88-tahun-2021>.
- Argimón S, Abudahab K, Goater RJE, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb Genom*. 2016;2(11):e000093. <https://doi.org/10.1099/mgen.0.000093>.
- Bayliss, S.C., Thorpe, H.A., Coyle, N.M., Sheppard, S.K. & Feil, E.J. (2019) PIRATE: A fast and scalable pangenomics toolbox for clustering diverged orthologues in bacteria. *GigaScience*. 8 (10), giz119. <https://doi.org/10.1093/gigascience/giz119>.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
- Carrera-Salinas A, González-Díaz A, Calatayud L, et al. Epidemiology and population structure of *Haemophilus influenzae* causing invasive disease. *Microb Genom*. 2021;7(12):000723. <https://doi.org/10.1099/mgen.0.000723>.
- Chen, L., Yang, J., Yu, J., Yao, Z., Sun, L., Shen, Y. & Jin, Q. (2005) VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Research*. 33 (Database issue), D325–328. <https://doi.org/10.1093/nar/gki008>.
- Clantin B, Hodak H, Willery E, Loch C, Jacob-Dubuisson F, Villeret V. The crystal structure of filamentous hemagglutinin secretion domain and its implications for the two-partner secretion pathway. *Proceedings of the National Academy of Sciences*. 2004;101(16):6194–9. <https://doi.org/10.1073/pnas.0400291101>.
- Cleland G, Leung C, Cheong Wan Sai, J., Francis, J., Heney, C. & Nourse, C. Paediatric invasive *Haemophilus influenzae* in Queensland, Australia, 2002–2011: Young Indigenous children remain at highest risk: Paediatric *Haemophilus influenzae* in Queensland. *Journal of Paediatrics and Child Health*. 2018;54(1):36–41. <https://doi.org/10.1111/jpc.13662>.
- Collins LF, Havers FP, Tunalı A, Thomas S, Clennon JA, Wiley Z, Tobin-D'Angelo M, Parrott T, Read TD, Satola SW, Petit RA, Farley MM. Invasive Nontypeable *Haemophilus influenzae* Infection Among Adults With HIV in Metropolitan Atlanta, Georgia, 2008–2018. *JAMA*. 2019;322(24):2399. <https://doi.org/10.1001/jama.2019.18800>.
- Davis GS, Patel M, Hammond J, Zhang L, Dawid S, Marrs CF, Gilsdorf JR. Prevalence, distribution, and sequence diversity of hmwA among commensal and otitis media non-typeable *Haemophilus influenzae*. *Infection, Genetics and Evolution*. 2014;28:223–32. <https://doi.org/10.1016/j.meegid.2014.09.035>.
- Didelot X, Wilson DJ. ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial Genomes. *PLOS Computational Biology*. 2015;11(2): e1004041. <https://doi.org/10.1371/journal.pcbi.1004041>.
- Efron A, Nápoli D, Neyro S, Juárez MDV, Moscoloni M, Eluchans NS, Regueira M, Lavayén S, Faccone D, Santos M. Laboratory surveillance of invasive *Haemophilus influenzae* disease in Argentina, 2011–2019. *Revista Argentina de Microbiología*. 2023;55(2):133–42. <https://doi.org/10.1016/j.ram.2022.08.002>.
- Eton V, Schroeter A, Kelly L, Kirlaw M, Tsang RSW, Ulanova M. Epidemiology of invasive pneumococcal and *Haemophilus influenzae* diseases in Northwestern Ontario, Canada, 2010–2015. *International Journal of Infectious Diseases*. 2017;65:27–33. <https://doi.org/10.1016/j.ijid.2017.09.016>.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29(8):1072–5. <https://doi.org/10.1093/bioinformatics/btt086>.
- Hasegawa Y, Arinuma Y, Tanaka S, Tono T, Tanaka T, Muramatsu T, Kondo J, Matsueda Y, Hoshiyama T, Wada T, Takayama Y, Yamaoka K. *Haemophilus influenzae* Non-type b Infection in an Adult Patient with Systemic Lupus Erythematosus. *Internal Medicine*. 2020;59(23):3097–101. <https://doi.org/10.2169/internalmedicine.4562-20>.
- Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, Wimalaratna H, Harrison OB, Sheppard SK, Cody AJ, Maiden MCJ. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology*. 2012;158(Pt 4):1005–15. <https://doi.org/10.1099/mic.0.055459-0>.
- Jolley, K.A., Bray, J.E. & Maiden, M.C.J. (2018) Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Research*. 3, 124. <https://doi.org/10.12688/wellcomeopenres.14826.1>.
- Khattak, Z. & Anjum, F. (2022) *Haemophilus influenzae* Infection. *StatPearls*. <https://www.statpearls.com/ArticleLibrary/viewarticle/91515>.
- Ladhani S, Slack MPE, Heath PT, von Gottberg A, Chandra M, Ramsay ME, Surveillance European Union Invasive Bacterial Infection, participants., Invasive *Haemophilus influenzae* Disease, Europe, 1996–2006. *Emerging Infectious Diseases*. 2010;16(3):455–63. <https://doi.org/10.3201/eid1603.090290>.
- Langereis JD, de Jonge MI. Invasive Disease Caused by Nontypeable *Haemophilus influenzae*. *Emerging Infectious Diseases*. 2015;21(10):1711–8. <https://doi.org/10.3201/eid2110.150004>.
- León ME, Kawabata A, Nagai M, Rojas L, Chamorro G, Zárate N, Gómez G, Leguizamón M, Irala J, Ortellado J, Franco R, Segovia N. Epidemiologic study of *Haemophilus influenzae* causing invasive and non-invasive disease in Paraguay (1999–2017). *Enfermedades infecciosas y microbiología clínica (English ed)*. 2021;39(2):59–64. <https://doi.org/10.1016/j.jemice.2020.11.014>.
- Liu B, Zheng D, Zhou S, Chen L, Yang J. VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Research*. 2022;50(D1):D912–7. <https://doi.org/10.1093/nar/gkab1107>.
- Maughan, H. & Redfield, R.J. (2009) Tracing the Evolution of Competence in *Haemophilus influenzae* I. Mokrousov (ed.). *PLoS ONE*. 4 (6), e5854. <https://doi.org/10.1371/journal.pone.0005854>.
- Mihara K, Tanabe T, Yamakawa Y, Funahashi T, Nakao H, Narimatsu S, Yamamoto S. Identification and transcriptional organization of a gene cluster involved in biosynthesis and transport of acinetobactin, a siderophore produced by *Acinetobacter baumannii* ATCC 19606T. *Microbiology*

- (Reading, England). 2004;150(Pt 8):2587–97. <https://doi.org/10.1099/mic.0.27141-0>.
27. National Institute for Communicable Diseases (2018) *GERMS South Africa: Annual Surveillance Review 2018*, p.50. <https://www.nicd.ac.za/wp-content/uploads/2019/11/GERMS-SA-AR-2018-Final.pdf>.
 28. Pimenta FC, Moiane B, Lessa FC, Venero A-KL, Moura I, Larson S, Massora S, Chaúque A, Tembe N, Mucavele H, Verani JR, Whitney CG, Sigauque B, Carvalho MGS. Dried blood spots for *Streptococcus pneumoniae* and *Haemophilus influenzae* detection and serotyping among children < 5 years old in rural Mozambique. *BMC Pediatrics*. 2020;20:326. <https://doi.org/10.1186/s12887-020-02209-3>.
 29. Pinto M, González-Díaz A, Machado MP, Duarte S, Vieira L, Carriço JA, Marti S, Bajanca-Lavado MP, Gomes JP. Insights into the population structure and pan-genome of *Haemophilus influenzae*. *Infection, Genetics and Evolution*. 2019;67:126–35. <https://doi.org/10.1016/j.meegid.2018.10.025>.
 30. Pribjelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo Assembler. *Current Protocols in Bioinformatics*. 2020;70(1): e102. <https://doi.org/10.1002/cpbi.102>.
 31. Rusmil K, Gunardi H, Fadlyana E, Soedjatmiko Dhamayanti, M., Sekartini, R., Satari, H.I., Risan, N.A., Prasetyo, D., Tarigan, R., Garheni, R., Milanti, M., Hadinegoro, S.R., Tanuwidjaja, S., Bachtar, N.S. & Sari, R.M. The immunogenicity, safety, and consistency of an Indonesia combined DTP-HB-Hib vaccine in expanded program on immunization schedule. *BMC Pediatrics*. 2015;15:219. <https://doi.org/10.1186/s12887-015-0525-2>.
 32. Safari, D., Lestari, A.N., Khoeri, M.M., Tafroji, W., Giri-Rachman, E.A., Harimurti, K. & Kurniati, N. (2020) Nasopharyngeal carriage and antimicrobial susceptibility profile of *Haemophilus influenzae* among patients infected with HIV in Jakarta, Indonesia. *Access Microbiology*. 2 (12), acmi000165. <https://doi.org/10.1099/acmi.0.000165>.
 33. Schürch AC, Arredondo-Alonso S, Willems RJA, Goering RV. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. *Clinical Microbiology and Infection*. 2018;24(4):350–4. <https://doi.org/10.1016/j.cmi.2017.12.016>.
 34. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30(14):2068–9. <https://doi.org/10.1093/bioinformatics/btu153>.
 35. Singh S, Bajorek B. Defining 'elderly' in clinical practice guidelines for pharmacotherapy. *Pharmacy Practice*. 2014;12(4):489. <https://doi.org/10.4321/s1886-36552014000400007>.
 36. Slack M, Esposito S, Haas H, Mihalyi A, Nissen M, Mukherjee P, Harrington L. *Haemophilus influenzae* type b disease in the era of conjugate vaccines: critical factors for successful eradication. *Expert Review of Vaccines*. 2020;19(10):903–17. <https://doi.org/10.1080/14760584.2020.1825948>.
 37. Slack MPE, Cripps AW, Grimwood K, Mackenzie GA, Ulanova M. Invasive *Haemophilus influenzae* Infections after 3 Decades of Hib Protein Conjugate Vaccine Use. *Clinical Microbiology Reviews*. 2021;34(3):e00028–21. <https://doi.org/10.1128/CMR.00028-21>.
 38. Slotved H-C, Johannesen TB, Stegger M, Fuursted K. Evaluation of molecular typing for national surveillance of invasive clinical *Haemophilus influenzae* isolates from Denmark. *Frontiers in Microbiology*. 2022;13:1030242. <https://doi.org/10.3389/fmicb.2022.1030242>.
 39. Soeters HM, Blain A, Pondo T, Doman B, Farley MM, Harrison LH, Lynfield R, Miller L, Petit S, Reingold A, Schaffner W, Thomas A, Zansky SM, Wang X, Briere EC. Current Epidemiology and Trends in Invasive *Haemophilus influenzae* Disease—United States, 2009–2015. *Clinical Infectious Diseases*. 2018;67(6):881–9. <https://doi.org/10.1093/cid/ciy187>.
 40. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312–3. <https://doi.org/10.1093/bioinformatics/btu033>.
 41. Staples M, Graham RMA, Jennison AV. Characterisation of invasive clinical *Haemophilus influenzae* isolates in Queensland, Australia using whole-genome sequencing. *Epidemiology and Infection*. 2017;145(8):1727–36. <https://doi.org/10.1017/S0950268817000450>.
 42. Suga S, Ishiwada N, Sasaki Y, Akeda H, Nishi J, Okada K, Fujieda M, Oda M, Asada K, Nakano T, Saitoh A, Hosoya M, Togashi T, Matsuoka M, Kimura K, Shibayama K. A nationwide population-based surveillance of invasive *Haemophilus influenzae* diseases in children after the introduction of the *Haemophilus influenzae* type b vaccine in Japan. *Vaccine*. 2018;36(38):5678–84. <https://doi.org/10.1016/j.vaccine.2018.08.029>.
 43. Wang X, Mair R, Hatcher C, Theodore MJ, Edmond K, Wu HM, Harcourt BH, Carvalho MDGS, Pimenta F, Nymadawa P, Altantsetseg D, Kirsch M, Satola SW, Cohn A, Messonnier NE, Mayer LW. Detection of bacterial pathogens in Mongolia meningitis surveillance with a new real-time PCR assay to detect *Haemophilus influenzae*. *International Journal of Medical Microbiology*. 2011;301(4):303–9. <https://doi.org/10.1016/j.ijmm.2010.11.004>.
 44. Watts SC, Holt KE. Hicap: In silico serotyping of the *Haemophilus influenzae* capsule locus. *J Clin Microbiol*. 2019;57:10.1128/jcm.00190-19. <https://doi.org/10.1128/jcm.00190-19>.
 45. Whittaker R, Economopoulou A, Dias JG, Bancroft E, Ramliden M, Celenz LP, Centre European, for Disease Prevention and Control Country Experts for Invasive *Haemophilus influenzae* Disease, Epidemiology of Invasive *Haemophilus influenzae* Disease, Europe, 2007–2014. *Emerging Infectious Diseases*. 2017;23(3):396–404. <https://doi.org/10.3201/eid2303.161552>.
 46. World Health Organization & Centers for Disease Control and Prevention (U.S.) (2011) *Laboratory Methods for Diagnosis of Meningitis Caused by Neisseria meningitidis, Streptococcus pneumoniae, Haemophilus influenzae: WHO Manual, 2nd ed.* pp.32–258.

Publisher's Note

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