Canadian Journal of Microbiology

Global population structure of *Haemophilus influenzae* serotype a (Hia) and emergence of invasive Hia disease: capsule switching or capsule replacement?

Journal:	<i>Canadian Journal of Microbiology</i>
Manuscript ID	cjm-2021-0152.R1
Manuscript Type:	Article
Date Submitted by the Author:	13-Jul-2021
Complete List of Authors:	Shuel, Michelle; National Microbiology Laboratory, Knox, Natalie; Public Health Agency of Canada, National Microbiology Laboratory; University of Manitoba, Department of Medical Microbiology and Infectious Diseases Tsang, Raymond S.W.; CNS Infection Division and Vaccine Preventable Bacterial Diseases Division,,
Keyword:	Population biology, <i>Haemophilus influenzae</i> , serotype a, capsule replacement
Is the invited manuscript for consideration in a Special Issue? :	Not applicable (regular submission)

SCHOLARONE™
Manuscripts

Global population structure of *Haemophilus influenzae* serotype a (Hia) and emergence of invasive Hia disease: capsule switching or capsule replacement?

Michelle Shuel¹, Natalie Knox^{2,3}, and Raymond S.W. Tsang^{1*}

Vaccine Preventable Bacterial Diseases¹, and Bioinformatics Section², National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, Manitoba, Canada; Department of Medical Microbiology and Infectious Diseases, University of Manitoba, Winnipeg, Manitoba, Canada³.

* Correspondence:

Raymond Tsang

National Microbiology Laboratory

1015 Arlington Street

Winnipeg, Manitoba, Canada R3E 3R2

Phone: 204-789-6020

Fax: 204-789-2018

Email: raymond.tsang@canada.ca

Abstract

The population structure of Hia was examined by interrogation of the *H. influenzae* MLST website. There were 196 entries of Hia with 55 sequence types (STs) identified (as of September 3, 2020). BURST analysis clustered related STs into four complexes with ST-23, ST-4, ST-21 and ST-62 identified as their ancestral STs. The majority of Hia entries (73.4%) and STs (65.5%) were identified as clonal division I (ST-23 and the ST-4 complexes). Only 43 (21.9%) entries and 14 STs (25.5%) were identified as clonal division II (ST-62 and ST-21 complexes). Current data suggested most invasive Hia belonged to clonal division I and the ST-23 complex while most clonal division II Hia were respiratory isolates with the exception of ST-62 which was common among invasive Hia in the U.S. southwest. Comparison of the capsule *bexABCD* genes from clonal divisions I and II strains showed sequence diversity with variations following the pattern of clonal divisions. Evidence from the literature and the current study suggests recent emergence of invasive Hia might be related to capsule replacement subsequent to the implementation of the Hib conjugate vaccine and possibly exacerbated by other conjugate vaccines that may have altered the microbial flora of the human respiratory tract.

Keywords: Population biology, *Haemophilus influenzae*, serotype a, capsule replacement

Introduction

Haemophilus influenzae is a Gram-negative coccobacillus that infects only the human host and normally colonizes the respiratory tract. From there it can either cause localized infections like otitis media and sinusitis or invade the blood stream to cause systemic infections like meningitis, pneumonia and sepsis (Slack 2006). One of the virulence factors of *H. influenzae* is the polysaccharide capsule, which also serves as the serotyping antigen. Six serotypes have been identified (a, b, c, d, e, and f), while strains without a capsule are termed non-typeable (Pittman 1931). The capsule (*cps*) locus, consisting of three regions, is responsible for synthesis of the polysaccharide capsule. Region II contains serotype-specific variable genes responsible for the synthesis of enzymes involved in assembling the capsular structure (Van Eldere et al. 1995, Satola et al. 2003). Regions I and III contain the common genes *bexABCD* and *hcsAB*, respectively, for the export or transport of the capsule structure to the cell surface and for post-translational modification of the capsule (Kroll et al. 1990, Sukupolvi-Petty et al. 2006). The capsule also serves as a protective antigen shielding the bacterium from the human host defence like phagocytosis and complement activation (Hallstrom et al. 2010). As the protective antigen, vaccines based on the capsule have been developed to target the most common serotype causing invasive infections (Kelly et al. 2004).

Before the introduction of the *H. influenzae* serotype b (Hib) conjugate vaccine, Hib was a common cause of childhood meningitis and other invasive diseases such as bacteremic pneumonia, septic arthritis and epiglottitis (Mason et al. 1982, Falla et al. 1993). Besides preventing invasive infections, the Hib conjugate vaccine also reduces or eliminates respiratory carriage of Hib and hence offers herd immunity (Mohle-Boetani et al. 1993, Barbour et al. 1995, Kelly et al. 2004). However, the Hib vaccine is serotype-specific and does not prevent infections

caused by other serotypes (i.e. a, c, d, e, and f) or non-typeable *H. influenzae* strains. Hence, after the introduction of the Hib conjugate vaccine in infant immunization programs, the epidemiology of invasive *H. influenzae* disease in both North America and Europe has changed dramatically (Whittaker et al. 2017, Soeters et al. 2018). In North America, Hib disease is now rare in children, but serotype a *H. influenzae* (Hia) has emerged to become a frequent cause of invasive infections among the Indigenous population (Tsang et al. 2014). Recent studies in the United States (US) have estimated that the overall incidence of invasive Hia disease in the US population has increased on average 13% annually from 2002 to 2015 (Soeters et al. 2018). Outbreaks of invasive Hia disease have also been reported in Alaska and Northern Canada (Hammitt et al. 2005, Bruce et al. 2013, Boisvert & Moore 2015, Tsang et al. 2016). In a U.S. population-based survey between 2011-2015, the severity of invasive Hia disease was evident with almost half (47.5%) and one quarter (25.3%) of the hospitalized patients either being admitted to intensive care units, and/or requiring mechanical ventilation, respectively (Bozio et al. 2020). Therefore, an Hia conjugate vaccine has been proposed as a priority in order to protect the vulnerable population (Boisvert & Moore 2015).

Population genetic studies on encapsulated *H. influenzae* have identified two clonal divisions: I and II, with most clinical isolates of Hia, Hib, serotype c, and serotype d belonging to clonal division I, while a small number of Hia and Hib, as well as all serotype f (Hif) belong to clonal division II (Musser et al. 1988). Serotype e (Hie) appears to be an intermediate between these two clonal divisions (Musser et al. 1988). Isolates of clonal division II as well as Hie are characterized by the presence of the copper/zinc superoxide-dismutase (*sodC*) gene (Kroll et al. 1991, Langford et al. 2002). A previous study based on multi-locus sequencing typing (MLST)

on invasive Hia in three Canadian provinces found four “clones” defined by their sequence types (STs): ST-4 as well as ST-23 and its related STs (single, double, and triple locus variants) were found to be negative for the *sodC* gene while ST-62 and ST-372 were positive for the *sodC* gene (Tsang et al. 2013), suggesting ST-23 and ST-4 belong to clonal division I, while ST-62 and ST-372 belong to clonal division II. A subsequent study by whole genome sequencing on a larger sample of Hia collected from across Canada confirmed our earlier findings on the overall population biology of Hia (Tsang et al. 2020).

Therefore, in order to understand the background on the emergence of invasive Hia disease, we examined the global population structure of Hia, examined the relationship of clonal types and disease, and explored reasons behind the recent increase in invasive Hia disease reported in North America.

Materials and Methods

Bacterial strains and genome sequence data.

Included in this study were 26 strains chosen to cover both clonal divisions as well as the most common serotypes of a, b, e, and f (Supplementary Table S1). There were 20 strains selected from our culture collection of both clinical isolates (collected from 2001 to 2019) and reference strains: 11 Hia (nine clinical isolates and two reference strains) from our previous study (Tsang et al. 2020), three Hib (one clinical isolate and two reference strains) and six Hif (five clinical isolates and one reference strain). Genome data for five strains, including three Hia, one Hib, and one Hif, together with nucleotide sequences of the *cps* genes from an Hie strain, were obtained from the public domain (Supplementary Table S1).

Multi-locus sequence typing (MLST) and interrogation of the *H. influenzae* MLST website.

MLST was carried out as previously described (Meats et al. 2003) and assignment of allele numbers and STs was done through the *H. influenzae* MLST website (<https://pubmlst.org/hinfluenzae/>). Traditional PCR and gene sequencing was used to determine STs for all isolates from our culture collection, while an *in silico* MLST analysis was performed on the publically available genomes by submitting the assembled contigs to the *H. influenzae* MLST website. To gain a broader perspective of Hia population structure across the world, we utilized the global resource of the *H. influenzae* MLST website, which had records for 5799 isolates and 2323 MLST profiles as of September 3, 2020. Isolates belonging to Hia in the MLST website were identified based on a search of the isolates database. A BURST diagram was generated for all Hia STs found in the database (as of September 3, 2020) using the online BURST analysis tool within the MLST website. The BURST program identified related STs as

single locus variant (SLV), double locus variant (DLV), and triple locus variants (TLV), as well as the ancestral or founding STs for each group. STs with identical alleles in at least four of their seven housekeeping genes were considered to form a clonal complex.

Whole genome sequencing and bioinformatics analysis

Genomic DNA was extracted from pure cultures using the Lucigen MasterPure Complete DNA and RNA Extraction Kit (Mandel Scientific, Guelph, ON) following the manufacturer's instructions. Multiplexed TruSeq DNA sequencing libraries (Illumina, San Diego, CA) were sequenced on an Illumina MiSeq (Illumina, San Diego, CA) as 2 x 300 bp paired-end reads. Whole genome sequence (WGS) data from each isolate was assembled using SPAdes (Bankevich et al. 2012).

The gene sequences for *bexABCD*, *hcsAB*, and *sodC* were identified based on the NCBI annotations in Hif strain KR494 (GenBank accession number CP005967) and used to BLAST [BLAST+ blastn] (Camacho et al. 2009, Cock et al. 2015) against WGS assemblies for Hia, Hib, and Hif isolates from our culture collection as well as the publically available *H. influenzae* genomes. The extracted sequences were aligned using software by DNASTAR Lasergene, which was also used to calculate sequence similarity (Madison, Wisconsin, USA). Construction of phylogenetic trees was performed on the phylogeny.fr platform using MUSCLE, Gblocks, PhyML, and TreeDyn for the alignment, curation, phylogeny, and tree rendering steps, respectively (Castresana 2000, Guindon & Gascuel 2003, Edgar 2004, Anisimova & Gascuel 2006, Dereeper et al. 2008, Dereeper et al. 2010). The *cps* region II genes (*acsI-4*, *bcsI-4* and *fcsI-3*) were extracted from the 13 of the 14 serotypes a, 4 serotype b, and 7 serotype f isolates

listed in Table S1 (isolates chosen were shown in Supplementary Figure S1), and compared in the same way as the other *cps* genes as described above.

Draft

Results and Discussion

Global population structure of Hia based on information in the *H. influenzae* MLST website,

As of September 3, 2020, there were 196 records and 55 STs associated with Hia isolates in the MLST website, including four records without a ST assigned due to incomplete allelic profiles.

The allelic profiles of the 55 STs associated with Hia are shown in Supplementary Table S2.

Figure 1 is a BURST diagram obtained from the *H. influenzae* MLST website that shows related Hia STs clustered together by having identical MLST gene alleles. The largest cluster of STs (32 STs) was grouped around ST-23 (the ancestral type or founder of the cluster as determined by BURST) with 17 STs being SLVs, 8 being DLVs, and 5 being TLVs of ST-23 as well as also one ST (ST-1755), which is a DLV of ST-403 but differed from the group founder ST-23 by four alleles (*atpG*, *fucK*, *mdh*, and *recA*). The second largest cluster consisted of eleven STs with ST-21 identified as the ancestral type. This cluster was found to have two sub-groups of four and seven STs, related through ST-21 and ST-372, with these two STs being TLVs of each other. Two smaller clusters made up of three or four STs were also identified, with ST-62 and ST-4 identified as their ancestral types, respectively. Besides these four clusters, there were five other STs (singletons) not related to any of these clusters nor to each other, including ST-6, ST-723, ST-747, ST-2117, and ST-2172. Isolates of ST-723, ST-747, ST-2117 and ST-2172 were from throat specimens associated with carriage, while ST-6 was a blood isolate.

When the *H. influenzae* MLST website was accessed in January 2017 (Tsang & Ulanova 2017), only 32 STs were found to be associated with Hia. Since then, the number of Hia-associated STs has increased to 55, an increase of 71.9%. This increase in the number of STs detected since 2017 was mainly driven by an increase in the number of STs in the ST-23 complex and the ST-

21/ST-372 complex, which has increased by 77.8% (from 18 STs to 32 STs) and 83.3% (from six STs to eleven STs), respectively. The size of the ST-4 complex has also increased by 100% from two STs to four STs. The increase in the number of Hia STs identified in the last 3.5 years (from January 2017 to September 2020) does not appear to reflect temporal increase in genetic diversity of Hia as many of the newer STs were identified based on historical isolates, which suggests better awareness of the importance of Hia and more research being dedicated to the genetic analysis of current and historical isolates. Of the 23 new STs identified in this study and since 2017, 13 (56.5%) were associated with Hia isolates recovered prior to 2017 (1981 to 2016), six STs were associated with isolates recovered in 2017 and three STs from 2018, while one ST was from an isolate without date of isolation.

Excluding the five singleton STs, and based on data from the *H. influenzae* MLST website, most (144 or 73.5%) entries and STs (36 or 65.5%) were associated with Hia isolates belonging to clonal division I (associated with ST-23 or ST-4 complexes), while only 43 (21.9%) entries and 14 (25.5%) STs were associated with Hia isolates that belonged to clonal division II (associated with ST-62 or ST-21/ST-372 complexes). Current information based on our own data (Tsang et al., 2013; 2019), data from the literature (Ulanova & Tsang 2014, Tsang & Ulanova 2017), as well as from the *H. influenzae* MLST website also suggests that most invasive Hia are associated with STs found in clonal division I (the ST-23 and ST-4 complexes), while the majority of isolates and STs associated with clonal division II were respiratory isolates from carriers and not involved in invasive disease with the exception of ST-62. Of the 18 ST-62 isolates described in the MLST website, four were associated with invasive disease based on their site of isolation (blood, CSF and joint), while the rest did not have information on source of isolation or disease

association. However, a study of invasive Hia infections in Utah children <18 years old in the United States between 2007 and 2017 found that 75% (21/28) of the case isolates were typed as ST-62 while the rest were typed as members of the ST-23 complex (Crandall et al. 2020). Two other Hia isolates in the United States (M21384 and SD2016_2) described by Watts & Holt (2019) and Terrat et al. (2020), respectively, were also identified as ST-62 in this study by *in-silico* analysis of their genomes (Figure 2 and Supplementary Table S1). M21384 was isolated in New Mexico (based on information in the *H. influenzae* MLST website) and SD2016_2 was isolated in California (Terrat et al. 2020). Besides M21384, two other ST-62 Hia (M21304 and M21387) found in the *H. influenzae* MLST website were also described to be isolated in New Mexico. This is in contrast to the absence of this ST among Hia isolates found in Alaska or Northern Canada (Tsang et al., 2017). Therefore, ST-62 Hia may appear to be more common in the American southwest, suggesting a potential geographical variation of Hia genotype in North America.

The overall population structure of Hia as well as the size and predominance of clonal division I isolates associated with invasive disease are very similar to what has been described for Hib (Musser et al. 1988), i.e. most invasive clinical isolates of Hib also belong to clonal division I (Musser et al. 1988). Indeed recent description of the clinical disease caused by Hia also confirmed their similarity to Hib disease (Bozio et al. 2020).

Capsule synthesizing genes, presence of *sodC* gene, and clonal structure of encapsulated *H. influenzae*.

The STs of the 25 genomes analyzed in this study were described in Figures 2 and 3, as well as Supplementary Table S1. No complete or draft genome data was available in the public domain for any Hie strain and only genetic data for the *cps* locus was available from NCBI GenBank and therefore, presence of *sodC* gene and ST could not be determined for the Hie data presented. The presence of the *sodC* gene was confirmed by an *in silico* analysis in all the Hif strains, all the ST-62 and the ST-372 Hia, as well as the clonal division II Hib strain NCTC 8468. The *sodC* gene was absent in the clonal division I Hia (ST-4 and ST-23 complexes) and clonal division I Hib strains (Eagan, 10810 and the clinical isolate B1).

To examine the relationship of capsule synthesis genes, serotypes, and clonal division, we performed an *in silico* analysis of *bexABCD* (region I genes) of the *cps* locus on all 25 genomes together with the Hie *cps* genes. Sequences of region I genes (*bexABCD*) showed distinct clustering based on clonal division with two major clusters found, which corresponded to clonal divisions I and II (Figure 2). Sub-clustering within the two main branches was found to be related to the serotypes of the isolates. The first major cluster consisted of clonal division I Hia and clonal division I Hib (upper half of Figure 2). Clonal division I Hia belonging to the ST-23 clonal complex (ST-23, ST-56, ST-576, ST-1290, and ST-1767) had only one single nucleotide polymorphism (SNP) among the six isolates. The *bexABCD* sequences of the ST-4 isolate when compared to that of the ST-23 clonal complex isolates showed 98.7% similarity. ST-4 and ST-23 are separate clones of clonal division I Hia as they do not share any common MLST housekeeping gene loci (Supplementary Table S2). The three clonal division I Hib strains showed identical *bexABCD* sequences and were related to the ST-23 clonal complex Hia by 96.7%. The second major cluster consisted of Hif, Hie and all clonal division II Hia and Hib

strains. Clonal division II Hia clustered closely with clonal division II Hib and showed 99.9% similarity, whereas clonal divisions I and II Hia were only 87.6% similar. All seven Hif strains had identical *bexABCD* sequences and showed 96.6% similarity to clonal division II Hia, but only 87.1% similarity to clonal division I Hia. Therefore, this more in-depth analysis of the *bexABCD* genes of encapsulated *H. influenzae* confirms our earlier finding based on partial *bexA* sequence (Zhou et al. 2007), which showed the genetic diversity of this gene appears to follow the division of encapsulated *H. influenzae* into the two clonal divisions identified by Musser et al. (1988). Comparison of *cps* region II genes from serotypes a, b, and f clearly separated them according to their serotypes regardless of their grouping by clonal divisions, thus confirming their unique sequences required for the biosynthesis of the structurally different serotype capsules (Supplementary Figure S1).

A phylogenetic tree created based on *hcsAB* (region III) sequences also split the 14 Hia and 4 Hib isolates (except strain 10810) into two groups based on the clonal divisions to which they belong (Figure 3). Similar to *bexABCD*, Hia isolates within clonal division I and clonal division II were highly related (99.8% and 99.7% similarity, respectively), but relatedness between Hia of the 2 clonal divisions were less similar at 94.7% similarity only.

Recent increase in *H. influenzae* serotype a strains associated with invasive disease was due to capsule replacement and not capsule switching.

The Hib conjugate vaccine offers protection against Hib by inducing functional antibodies to the polyribosyl ribitol phosphate (PRP) capsule which is specific for Hib (Schreiber et al. 1986). The anti-PRP antibodies are bactericidal through antibody-mediated opsonophagocytosis and

complement-dependent bacterial lysis (Moxon et al. 1988). Since there are six serotypes of *H. influenzae* as well as non-encapsulated strains, vaccine induced anti-PRP antibodies may exert pressure on the *H. influenzae* population to change and adapt to the vaccine pressure. In order to evade the vaccine pressure, the *H. influenzae* may change through two mechanisms that involve their serotype capsule antigens: capsule or serotype switching and capsule or serotype replacement (Lipsitch 1999, Tsang 2007). In capsule switching, the Hib strain may recombine with a non-Hib strain to swap their *cps* genes in order to synthesize a non-Hib capsule, such as a Hia capsule, to evade the Hib vaccine induced anti-PRP antibodies. In capsule replacement, strains with a vaccine capsule type such as Hib that used to be prevalent in the nasopharynx prior to vaccine use have been reduced or eliminated by the Hib vaccine induced immunity (Mohle-Boetani et al. 1993, Barbour et al. 1995, Kelly et al. 2004) and are being replaced by strains with non-vaccine capsule types such as Hia. If Hia has the virulence potential as reported only second to Hib in order of virulence by animal experimental study (Zwahlen et al. 1989), then the emergence of Hia to cause invasive disease and replaces the disease that used to be caused by Hib may become a predictable outcome.

During our current interrogation of the *H. influenzae* MLST website for Hia strains, 1 blood isolate (id 3449) was identified as ST-6, which is a common ST of Hib strains and the strain was described as serotype a but genotype b suggesting an error or inconsistency in the serotype results. Another throat isolate (id 1351) identified as serotype a and ST-747 (in the MLST website), was found to have 6 MLST housekeeping gene alleles identical to another throat isolate (isolated from the same year and same country) identified as serotype f and ST-748 (id 1352). Five of the MLST housekeeping gene alleles in ST-747 (*adh*, *atpG*, *frdB*, *fucK*, and *recA*) are

commonly found in STs associated with Hif strains (as observed in the *H. influenzae* MLST database). However, ST-747 has not been found in any Hif that we have in our culture collection ($n = 260$ isolates) or in the *H. influenzae* MLST website (number of Hif found on the MLST website was 266 as of September 3, 2020). Therefore, these data taken together suggest that capsule or serotype switching in *H. influenzae* is uncommon. This further confirms our previous analysis of the *H. influenzae* MLST website (Tsang 2008). Then what has caused the recent emergence of Hia infections in North America?

A recent editorial has suggested that capsule or serotype replacement is a likely cause of the emergence of invasive Hia disease in the past decade (Ulanova 2020). As the respiratory flora changes with the introduction of a number of powerful conjugate vaccines against the previously dominant organisms such as Hib, some pneumococcal serotypes found in the pneumococcal seven valent, ten valent, or thirteen valent conjugate vaccines (PCV-7, PCV-10, or PCV-13), as well as the meningococcal serogroups A, C, Y, and W, the ecological niches that used to be occupied by these pathogens will be replaced by serotypes or serogroups that are not covered by the vaccines. Once these newer serotypes colonize the respiratory tract, they may have the potential to cause disease especially if they have the virulence potential. Indeed, virulence in Hia has been shown to be only second to Hib when experimental animals were infected with isogenic mutants expressing different capsular serotypes (Zwahlen et al. 1989). Hia carriage is very common (up to 61% for those under the age of ten) among contacts of invasive Hia cases (Nolen et al. 2020). Carriage of Hib in children appears to be the major reservoir of infection (Oh et al. 2008). Also similar to Hib, it is possible that increase in disease is preceded by an increase in carriage detection (Jacups et al. 2011). Therefore, conjugate vaccines that can eliminate

oropharyngeal carriage would be an ideal tool for control of invasive Hia disease in the vulnerable population.

Based on the experience with introduction of PCVs and the subsequent changes in the pneumococcal disease pattern in the US with replacement disease first observed in Alaskan native children (Peters & Poehling 2007), could the significant increase in invasive Hia disease in the Indigenous population in Alaska and Northern Canada be an early warning signal for a potentially similar increase in invasive Hia disease in the general population? Pneumococcal disease due to serotype replacement resulted in significantly higher rates of disease in Alaska native children (244.6 cases per 100,000 children < 2 years old for 2004-2006) than in the general US population as recorded by US CDC's ABC surveillance (19.9 cases per 100,000 children < 5 years old in 2004) (Singleton et al. 2007, Hick et al. 2007). If experience in pneumococcal disease is a beacon of what may happen to invasive *H. influenzae* disease, then a better surveillance system and vaccine preparation (Cox et al. 2017) may be required to deal with the changing epidemiology of invasive *H. influenzae* disease.

Conclusion

Our current knowledge on invasive Hia disease suggests most disease-associated isolates belong to clonal division I, especially those related to the ST-23 complex. Minor clones such as ST-4 in clonal division I and ST-62 in clonal division II are also involved in causing invasive disease. Geographical variations in the clonal types involved in invasive disease appears to happen, e.g. ST-62 in the American southwest (Crandall et al. 2020) and ST-4 in Brazil (Lima et al. 2010). The significance of the ST-23 clone in causing the majority of invasive Hia disease is probably

reflected by the large number of genetic variants (32 different STs) of the ST-23 complex documented in the last two decades. The increases in invasive Hia disease is likely a result of capsule or serotype replacement with ecological niches previously occupied by Hib and vacated by the Hib conjugate vaccine thereby allowing Hia to colonize and cause diseases. The role of other conjugate vaccines against common respiratory tract pathogens in altering the respiratory flora has also been suggested as a possible contributing factor (Ulanova 2020).

Declarations

Opinions expressed in this manuscript are those of the authors and they do not represent the view of the National Microbiology Laboratory nor the Public Health Agency of Canada.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in Genbank or the MLST website:

NMLHia1 (<https://www.ncbi.nlm.nih.gov/nuccore/CP017811>)

M21384 (<https://www.ncbi.nlm.nih.gov/nuccore/CP031250>)

SD2016_2 (https://www.ncbi.nlm.nih.gov/nuccore/NZ_SWJY01000004)

10810 (<https://www.ncbi.nlm.nih.gov/nuccore/FQ312006>)

KR494 (<https://www.ncbi.nlm.nih.gov/nuccore/CP005967>)

Hie *cps* operon (<https://www.ncbi.nlm.nih.gov/nuccore/FM882247>),

LNP31006:

https://pubmlst.org/bigssdb?db=pubmlst_hinfluenzae_isolates&page=seqbin&isolate_id=5609

NML WGS Data: available upon request.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by funding from the Public Health Agency of Canada as part of the routine laboratory surveillance of invasive *Haemophilus influenzae* disease.

Authors' contributions

Conceptualization: R.S.W.T.; Methodology: R.S.W.T., M.S., N.K. Data analysis: R.S.W.T., M.S. Writing of original draft: R.S.W.T., M.S.; Review and editing: All.

Acknowledgements

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.

The authors wish to thank the Genomics Core Facility at the National Microbiology Laboratory (NML) for the sequencing work. We also thank the Directors and staff of the Provincial and Territorial Public Health Laboratories for their support of the NML's laboratory surveillance program of invasive *H. influenzae* disease.

References

- Anisimova, M., and Gascuel, O. 2006. Approximate likelihood ratio test for branches: A fast, accurate and powerful alternative. *Systematic Biology*, **55**(4): 539-552.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, **19**(5): 455-477.
- Barbour, M.L., Mayon-White, R.T., Coles, C., Crook, D.W., and Moxon, E.R. 1995. The impact of conjugate vaccine on carriage of *Haemophilus influenzae* type b. *The Journal of Infectious Diseases*, **171**(1): 93-98.
- Boisvert, A.A., and Moore, D. 2015. Invasive disease due to *Haemophilus influenzae* type a in children in Canada's North: a priority for prevention. *Canadian Journal of Infectious Diseases and Medical Microbiology*, **26**(6): 291-292.
- Bozio, C.H., Blain, A., Edge, K., Farley, M.M., Harrison, L.H., Poissant, T., et al. 2020. Clinical characteristics and adverse clinical outcome of invasive *Haemophilus influenzae* serotype a cases – United States, 2011-2015. *Clinical Infectious Diseases*, <https://www.doi.org/10.1093/cid/ciaa990>.
- Bruce, M.G., Zulz, T., DeByle, C., Singleton, R., Hurlburt, D., Bruden, D., et al. 2013. *Haemophilus influenzae* serotype a invasive disease, Alaska, USA, 1983-2011. *Emerging Infectious Diseases*, **19**(6): 932-937.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics*, **10**: 421. <https://doi.org/10.1186/1471-2105-10-421>
- Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, **17**(4): 540-552.

- Chevenet, F., Brun, C., Banuls, A.L., Jacq, B., Chisten, R. 2006. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics*, **10**(7): 439.
- Cock, P.A.J., Chilton, J.M., Gruning, B., Johnson, J.E., and Soranzo, N. 2015. NCBI BLAST+ integrated into Galaxy. *Gigascience*, **4**: 39.
- Cox, A.D., Barreto, L., Ulanova, M., Bruce, M.G., and Tsang, R.S.W. 2017. The emerging *Haemophilus influenzae* serotype a infection and a potential vaccine: implementation science in action. *Canadian Communicable Disease Report*, **43**: 85-88.
- Crandall, H., Christiansen, J., Varghese, A.A., Russon, A., Korgenski, E.K., Bengtson, E.K., et al. 2020. Clinical and molecular epidemiology of invasive *Haemophilus influenzae* serotype a infections in Utah children. *Journal of the Pediatric Infectious Diseases Society*, **9**(6):650-655.
- Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., et al. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research*, **36**(2): W465-9.
- Dereeper, A., Audic, S., Claverie, J.M., Blanc, G. 2010. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC Evolutionary Biology*, **10**(8): [https://www.doi: 10.1186/1471-2148-10-8](https://www.doi.org/10.1186/1471-2148-10-8).
- Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**(5): 1792-1797.
- Falla, T.J., Dobson, S.R.M., Crook, D.W.M., Kraak, W.A., Nichols, W.W., Anderson, E.C., et al. 1993. Population-based study of non-typeable *Haemophilus influenzae* invasive disease in children and neonates. *Lancet* **341**: 851-854.
- Guindon, S., and Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**(5): 696-704.

- Hallstrom, T., and Riesbeck, K. 2010. *Haemophilus influenzae* and the complement system. Trends in Microbiology, **18**(6): 258-265.
- Hammit, L.L., Block, S., Hennessy, T.W., Debye, C., Peters, H., Parkinson, A., et al. 2005. Outbreak of invasive *Haemophilus influenzae* serotype a disease. The Pediatric Infectious Disease Journal, **24**(5): 453-456.
- Hicks, L.A., Harrison, L.H., Flannery, B., Hadler, J.L., Schaffner, W., Craig, A.S., et al. 2007. Pneumococcal disease due to non-pneumococcal conjugate vaccine (PCV7) serotypes in the United States during the era of widespread PCV7 vaccination, 1998-2004. The Journal of Infectious Diseases, **196** (9): 1346-1354.
- Jacups, S.P. 2011. The continuing role of *Haemophilus influenzae* type b carriage surveillance as a mechanism for early detection of invasive disease activity. Human Vaccines, **7**(12): 1254-1260.
- Kelly, D.F., Moxon, E.R., and Pollard, A.J. 2004. *Haemophilus influenzae* type b conjugate vaccines. Immunology, **113**(2): 163-174.
- Kroll, J.S., Loynds, B., Brophy, L.N., and Moxon, E.R. 1990. The *bex* locus in encapsulated *Haemophilus influenzae*: a chromosomal region involved in capsule polysaccharide export. Molecular Microbiology, **4**(11): 1853-1862.
- Kroll, J.S., Langford, P.R., and Loynds, B.M. 1991. Copper-zinc superoxide dismutase of *Haemophilus influenzae* and *H. parainfluenzae*. Journal of Bacteriology, **173**(23): 7449-7457.
- Langford, P.R., Sheehan, B.J., Shaikh, T., and Kroll, J.S. 2002. Active copper- and zinc-containing superoxide dismutase in the cryptic genospecies of *Haemophilus* causing urogenital and neonatal infections discriminates them from *Haemophilus influenzae sensu stricto*. Journal of Clinical Microbiology, **40**(1): 268-270.

- Lima, J.B., Ribeiro, G.S., Cordeiro, S.M., Gouveia, E.L., Salgado, K., Spratt, B.G., et al. 2010. Poor clinical outcome for meningitis caused by *Haemophilus influenzae* serotype a strains containing the IS1016-*bexA* deletion. *The Journal of Infectious Diseases*, **202**(10): 1577-1584.
- Lipsitch, M. 1999. Bacterial vaccines and serotype replacement: Lessons from *Haemophilus influenzae* and prospects for *Streptococcus pneumoniae*. *Emerging Infectious Diseases*, **5**(3): 336-345.
- Mason, E.O., Kaplan, S.L., Lamberthm, L.B., Hinds, D.B., Kvernland, S.J., Loiselle, E.M., et al. 1982. Serotype and ampicillin susceptibility of *Haemophilus influenzae* causing systemic infections in children: 3 years of experience. *Journal of Clinical Microbiology*, **15**(4): 543-546.
- Meats, E., Feil, E.J., Stringer, S., Cody, A.J., Goldstein, R., Kroll, J.S., et al. 2003. Characterization of encapsulated and nonencapsulated *Haemophilus influenzae* and determination of phylogenetic relationships by multilocus sequence typing. *Journal of Clinical Microbiology*, **41**(4): 1623-1636.
- Mohle-Boetani, J.C., Ajello, G., Brenman, E., Deaver, K.A., Harvey, C., Plikaytis, B.D., et al. 1993. Carriage of *Haemophilus influenzae* type b in children after widespread vaccination with conjugate *Haemophilus influenzae* type b vaccines. *The Pediatric Infectious Disease Journal*, **12**(7): 589-593.
- Moxon, E.R. and Kroll, J.S. 1988. Type b capsular polysaccharide as a virulence factor of *Haemophilus influenzae*. *Vaccine*, **6**(2): 113-115.
- Musser, J.M., Kroll, J.S., Moxon, E.R., and Selander, R.K. 1988. Clonal population structure of encapsulated *Haemophilus influenzae*. *Infection and Immunity*, **56**(8): 1837-1845.

- Musser, J.M., Kroll, J.S., Granoff, D.M., Moxon, E.R., Brodeur, B.R., Campos, J., et al. 1990. Global genetic structure and molecular epidemiology of encapsulated *Haemophilus influenzae*. *Reviews of infectious diseases*, **12**(1): 75-111.
- Nolen, L.D., Tiffany, A., DeByle, C., Bruden, D., Thompson, G., Reasonover, A., et al. 2020. *Haemophilus influenzae* serotype a (Hia) carriage in a small Alaska community after a cluster of invasive Hia disease, 2018. *Clinical Infectious Diseases*, **73**(2):e280-e286.
- Oh, S.Y., Griffiths, D., John, T., Lee, C.Y., Yu, L.M., McCarthy, N., et al. 2008. School-aged children: a reservoir for continued circulation of *Haemophilus influenzae* type b in the United Kingdom. *The Journal of Infectious Disease*, **197**(9): 1275-1281.
- Peters, T.R., and Poehling, K.A. 2007. Invasive pneumococcal disease: the target is moving. *The Journal of the American Medical Association*, **297**(16): 1825-1826.
- Pittman, M. 1931. Variations and type specificity in the bacterial species *Haemophilus influenzae*. *Journal of Experimental Medicine*, **53**(4): 471-492.
- Satola, S.W., Schirmer, P.L., and Farley, M.M. 2003. Genetic analysis of the capsule locus of *Haemophilus influenzae* serotype f. *Infection and Immunity*, **71**(12): 7202-7207.
- Schreiber, J.R., Barrus, V., Cates, K.L., and Siber, G.R. 1986. Functional characterization of human IgG, IgM, and IgA antibody directed to the capsule of *Haemophilus influenzae* type b. *The Journal of Infectious Diseases* **153**(1): 8-16.
- Singleton, R.J., Hennessy, T.W., Bulkow, L.R., Hammitt, L.L., Zulz, T., Hurlburt, D.A., et al. 2007. Invasive pneumococcal disease caused by non vaccine serotypes among Alaska native children with high levels of 7-valent pneumococcal conjugate vaccine coverage. *The Journal of American Medical Association*, **297**(16): 1784-1792.

Slack, M.P.E. *Haemophilus*. In: Borriello SP, Murray PR, Funke G, editors. Topley & Wilson's microbiology and microbial infections. 10th ed., Vol 2. London, UK: Hodder Arnold, 2006, pages 1692-1718.

Soeters, H.M., Blain, A., Pondo, T., Doman, B., Farley, M.M., Harrison, L.H., et al. 2018. Current epidemiology and trends in invasive *Haemophilus influenzae* disease – United States, 2009-2015. Clinical Infectious Diseases, **67**(6): 881-889.

Soeters, H.M., Oliver, S.E., Plumb, I.D., Blain, A.E., Zulz, T., Simons, B.C., et al. 2020. Epidemiology of invasive *Haemophilus influenzae* serotype a disease – United States, 2008-2017. Clinical Infectious Diseases, **73**(2):e371-e379.

Sukupolvi-Petty, S., Grass, S., St. Geme III, J.W. 2006. The *Haemophilus influenzae* type b *hcsA* and *hcsB* gene products facilitate transport of capsule polysaccharide across the outer membrane and are essential for virulence. Journal of Bacteriology, **188**(11): 3870-3877.

Terrat, Y., Farnaes, L., Bradley, J., Thomas, N., and Shapiro, B.J. 2020. Two cases of type-a *Haemophilus influenzae* within the same week in the same hospital are phylogenetically unrelated but recently exchanged capsule genes. Microbial Genomics, **6**(4): e000348.

Tsang, R.S.W. 2007. Capsule switching and capsule replacement in vaccine-preventable bacterial diseases. Lancet Infectious Diseases, **7**(9): 569-570.

Tsang, R.S.W. 2008. Serotyping and population genetics of invasive *Haemophilus influenzae*. J Journal of Clinical Microbiology, **46**(3):1159.

Tsang, R.S.W., Shuel, M., Wylie, J., Lefebvre, B., Hoang, L., and Law, D.K.S. 2013. Population genetics of *Haemophilus influenzae* serotype a in three Canadian provinces. Canadian Journal of Microbiology, **59**(5): 362-364.

Tsang, R.S.W., Bruce, M.G., Lem, M., Barreto, L., and Ulanova, M. 2014. A review of invasive *Haemophilus influenzae* disease in the Indigenous populations of North America. *Epidemiology and Infection*, **142**(7): 1344-1354.

Tsang, R.S.W., Li, Y.A., Mullen, A., Baikie, M., Whyte, K., Shuel, M., et al. 2016. Laboratory characterization of invasive *Haemophilus influenzae* isolates from Nunavut, Canada, 2000-2012. *International Journal of Circumpolar Health*, **75**: 29798.

Tsang, R.S.W., Proulx, J.F., Hayden, K., Shuel, M., Lefebvre, B., Boisvert, A.A., et al. 2017. Characteristics of invasive *Haemophilus influenzae* serotype a (Hia) from Nunavik, Canada and comparison with Hia strains in other North American Arctic regions. *International Journal of Infectious Diseases*, **57**: 104-107.

Tsang, R.S.W., and Ulanova, M. 2017. The changing epidemiology of invasive *Haemophilus influenzae* disease: Emergence and global presence of serotype a strains that may require a new vaccine for control. *Vaccine*, **35**: 4270-4275.

Tsang, R.S.W., Shuel, M., Ahmad, T., Hayden, K., Knox, N., Van Domselaar, G., et al. 2020. Whole genome sequencing to study the phylogenetic structure of serotype a *Haemophilus influenzae* recovered from patients in Canada. *Canadian Journal of Microbiology*, **66**(2): 99-110.

Ulanova, M., and Tsang, R.S.W. 2014. *Haemophilus influenzae* serotype a as a cause of serious invasive infections. *Lancet Infectious Diseases*, **14**(1):70-82.

Ulanova, M. 2020. Invasive *Haemophilus influenzae* serotype a disease in the Hib conjugate vaccine era: where are we going? *Clinical Infectious Diseases*, **73**(2):e380-e382.

Van Eldere, J., Brophy, L., Loynds, B., Celis, P., Hannock, I., Carman, S., et al. 1995. Region II of the *Haemophilus influenzae* type b capsulation locus is involved in serotype-specific polysaccharide synthesis. *Molecular Microbiology*, **15**(1): 107-118.

- Watts, S.C., and Holt, K.E. 2019. hicap: *In Silico* serotyping of the *Haemophilus influenzae* capsule locus. *Journal of Clinical Microbiology*, **57**(6): e00190-19.
- Whittaker, R., Economopoulou, A., Gomez Diaz, J., Bancroft, E., Ramliden, M., Celentano, L.P., European Center for Disease Prevention and Control Country Experts for invasive *Haemophilus influenzae* disease. 2017. Epidemiology of invasive *Haemophilus influenzae* disease, Europe, 2007-2014. *Emerging Infectious Diseases* **23**(3): 396-404.
- Zhou, J., Law, D.K.S., Sill, M.L., and Tsang, R.S.W. 2007. Nucleotide sequence diversity of the *bexA* gene in serotypeable *Haemophilus influenzae* strains recovered from invasive disease patients in Canada. *Journal of Clinical Microbiology*, **45**(6): 1996-1999.
- Zwahlen, A., Kroll, J.S., Rubin, L.G., and Moxon, E.R. 1989. The molecular basis of pathogenicity in *Haemophilus influenzae*: comparative virulence of genetically-related capsular transformants and correlation with changes at the capsulation locus *cap*. *Microbial Pathogenesis*, **7**(3): 225-235.

Figure 1. BURST diagram to show the global population biology of *Haemophilus influenzae* serotype a (Hia) based on sequence types (STs) identified in the *H. influenzae* MLST database (<https://pubmlst.org/hinfluenzae/>). The ST-21 and ST-372 clusters are related and are triple locus variants of each other. ST-60 and ST-2028 (not displayed by the BURST diagram) are triple locus variants of ST-2304. ST-2304 is a single locus variant of ST-372 and a triple locus variant of ST-21. Also not shown in the diagram are singletons, which are not related to each other or any other ST, and they include ST-6, ST-723, ST-747, ST-2117, and ST-2172.

Figure 2. Phylogenetic tree created using the phylogeny.fr platform for *bexABCD* sequences from 26 *Haemophilus influenzae* isolates, made up of 7 serotype a (Hia) and 3 serotype b (Hib) isolates representing clonal division I (CDI); 7 Hia, 1 Hib, and 7 serotype f (Hif) isolates representing clonal division II (CDII); and 1 serotype e (Hie) isolate. Isolates were designated by a prefix representing their serotype followed by the isolate or strain number, sequence type (ST), and clonal division. The scale bar represents 30% sequence divergence.

Figure 3. Phylogenetic tree created using the phylogeny.fr platform for *hcsAB* sequences from 26 *Haemophilus influenzae* isolates, made up of 7 serotype a (Hia) and 3 serotype b (Hib) isolates representing clonal division I (CDI); 7 Hia, 1 Hib, and 7 serotype f (Hif) isolates representing clonal division II (CDII); and 1 serotype e (Hie) isolate. Isolates were designated by a prefix representing their serotype followed by the isolate or strain number, sequence type (ST), and clonal division. The scale bar represents 30% sequence divergence.

Figure 2. Phylogenetic tree created using the phylogeny.fr platform for *bexABCD* sequences from 26 *Haemophilus influenzae* isolates, made up of 7 serotype a (Hia) and 3 serotype b (Hib) isolates representing clonal division I (CDI); 7 Hia, 1 Hib, and 7 serotype f (Hif) isolates representing clonal division II (CDII); and 1 serotype e (Hie) isolate. Isolates were designated by a prefix representing their serotype followed by the isolate or strain number, sequence type (ST), and clonal division. The scale bar represents 30% sequence divergence.

Figure 3. Phylogenetic tree created using the phylogeny.fr platform for *hcsAB* sequences from 26 *Haemophilus influenzae* isolates, made up of 7 serotype a (Hia) and 3 serotype b (Hib) isolates representing clonal division I (CDI); 7 Hia, 1 Hib, and 7 serotype f (Hif) isolates representing clonal division II (CDII); and 1 serotype e (Hie) isolate. Isolates were designated by a prefix representing their serotype followed by the isolate or strain number, sequence type (ST), and clonal division. The scale bar represents 30% sequence divergence.