Analysis of invasive non-typeable *Haemophilus influenzae* isolates reveals a selection for the expression state of particular phase-variable lipooligosaccharide biosynthetic genes.

Zachary N. Phillips\(^1\), Charles Brizuela\(^1\), Amy V. Jennison\(^2\), Megan Staples\(^2\), Keith Grimwood\(^3\), Kate L. Seib\(^1\), Michael P. Jennings\(^1\), and John M. Atack\(^1\)

\(^1\) Institute for Glycomics, Griffith University, Gold Coast, Queensland, Australia

\(^2\) Queensland Department of Health, Public Health Microbiology, Forensic and Scientific Services, Brisbane, Queensland, Australia

\(^3\) Menzies Health Institute Queensland, Griffith University, and Department of Infectious Diseases, Gold Coast Health, Queensland, Australia

Address for correspondence - John M Atack (JMA) or Michael P Jennings (MPJ); Institute for Glycomics, Griffith University, Gold Coast, QLD 4215, Australia. Tel: +61 (7) 555 80580 (JMA) or +61 (7) 555 27050 (MPJ); Email: j.atack@griffith.edu.au (JMA) or m.jennings@griffith.edu.au (MPJ)

Keywords: NTHi, invasive disease, LOS, phase-variation, glycosyltransferase
Abstract

Non-typeable *Haemophilus influenzae* (NTHi) is a major human pathogen, responsible for several acute and chronic infections of the respiratory tract. The incidence of invasive infections caused by NTHi is increasing world-wide. NTHi is able to colonise the nasopharynx asymptomatically, and the exact change(s) responsible for transition from benign carriage to overt disease are not understood. We have previously reported that phase-variation (the rapid and reversible ON-OFF switching of gene expression) of particular lipooligosaccharide (LOS) glycosyltransferases occurs during transition from colonising the nasopharynx to invading the middle ear. Variation in the structure of the LOS is dependent on the ON/OFF expression status of each of the glycosyltransferases responsible for LOS biosynthesis. In this study we surveyed a collection of invasive NTHi isolates for ON/OFF expression status of seven phase-variable LOS glycosyltransferases. We report that the expression state of the LOS biosynthetic genes *oafA* ON and *lic2A* OFF show a correlation with invasive NTHi isolates. We hypothesise that these gene expression changes contribute to the invasive potential of NTHi. OafA expression, which is responsible for the addition of an O-acetyl group onto the LOS, has been shown to impart a phenotype of increased serum resistance, and may serve as a marker for invasive NTHi.
Introduction

Non-typeable Haemophilus influenzae (NTHi) is a clinically significant bacterial pathogen of global relevance. NTHi is able to colonise the human nasopharynx asymptomatically, but is also responsible for acute and chronic infections of the respiratory tract, including middle ear infection (otitis media) in children (1), acute exacerbations in protracted bacterial bronchitis, chronic obstructive pulmonary disease and bronchiectasis (2, 3), and community-acquired pneumonia in adults (4). Since the introduction of a vaccine against H. influenzae serotype b (Hib), the incidence of invasive infection caused by NTHi has increased significantly worldwide (5, 6). NTHi is now a major cause of severe invasive disease in neonates, and responsible for invasive infections in children that have significant comorbidities (7, 8). NTHi invasive infections are fatal in ~10% of children between 2 and 4 years old, and ~17% in children under the age of 1 (9, 10). The increase in invasive disease caused by NTHi is likely due to multiple factors, including increasing numbers of vulnerable patient populations with complex comorbidities, rather than simply Hib vaccine-induced strain replacement (5).

Financial and pathological burdens of NTHi are increasing annually in the absence of an NTHi vaccine and amplified by emerging antibiotic resistant strains (11, 12). Several studies have investigated potential associations between the expression of certain virulence factors and invasive NTHi isolates (8, 13, 14), but none proved conclusive in demonstrating a link between any particular factor and the invasiveness of NTHi.

Phase-variation is the random and reversible switching of gene expression (15). Phase-variable gene expression can occur by several mechanisms, including homologous recombination between allelic variants, or variation in the length of simple sequence repeats (SSRs) (15). Phase-variation mediated by slipped-strand mispairing of SSRs located within, or associated with, an open-reading frame (ORF) commonly leads to the biphasic ‘ON-OFF’ switching of gene expression (15). This results in the encoded protein either being expressed
transcriptional termination is introduced (15). The length of SSR tracts has been shown to correlate with rates of phase-variation (16-18), with longer tracts exhibiting higher rates of phase-variation. The ability to produce multiple phenotypic variants within a bacterial population promotes strain adaptability and survival, and allows bacteria to evade host immune responses (15). Lipooligosaccharide (LOS) is a major NTHi virulence factor, and LOS presence has been shown to contribute to survival in-vivo (19, 20). Many NTHi LOS biosynthetic genes contain SSR tracts and are phase-variably expressed (21, 22). Phase-variable LOS biosynthetic genes include lic1A, encoding a phosphorylcholine transferase (23), lic2A encoding a galactosyltransferase (24), lic3A and lic3B encoding related sialyltransferases (20, 25), lex2A encoding a glucosyltransferase (26), lgtC encoding a galactosyltransferase (27), and oafA encoding an O-acetyltransferase (28) (a summary of NTHi LOS is presented in Figure 1A). Therefore ON/OFF switching of the expression of these glycosyltransferases will result in different LOS structures within an NTHi population.

Based on previous findings, and the importance of LOS in NTHi pathobiology, we hypothesised that the expression of individual LOS biosynthetic gene loci may be present or absent, or a particular expression status may be selected for (phase-varied ON) or against (phase-varied OFF) during invasive NTHi infection. We used two extensive, unique collections of NTHi taken in South-East Queensland, Australia: one containing invasive NTHi isolates collected over twenty years (29), and the second containing nasal swabs from healthy children over the first two years of life, the ORChID collection (30). By comparing isolates from the invasive collection to those in the carriage collection, we were able to
investigate if differences in LOS structure occurred during invasive disease compared to carriage. We demonstrate that the expression status of particular LOS biosynthetic genes (lic2A and oafA) appear to be selected for in invasive NTHi isolates compared to NTHi carriage isolates.

Results

By using our fluorescent PCR approach coupled to fragment length analysis, we have been able to determine the ON-OFF expression status of each of seven phase-variable LOS biosynthetic genes (lic1A, lic2A, lic3A, lic3B, lex2A, lgtC, oafA: see Figure 1) in 70 invasive NTHi isolates collected in South-East Queensland, Australia (29). Where PCR products could not be produced for individual genes despite multiple attempts, we analysed the genome sequences present for invasive isolates (accession number PRJEB18702) to confirm that these genes were in-fact not present in those particular isolates (data not shown). In previous studies of this type, it has also been demonstrated that not all strains contain all seven LOS biosynthetic loci (19). By comparing the ON/OFF expression status of these genes in invasive isolates to NTHi carriage isolates from the same region (30), we were able to determine if particular genes are selected for during NTHi invasive infections. Our results show that five of these genes, lic1A, lic3A, lic3B, lex2A, and lgtC, demonstrated no statistically significant difference for either an ON or an OFF expression state in invasive isolates, nor showed a significant difference from the ON-OFF status of carriage isolates. All data from fragment length analysis is presented in Supplementary Data 1.

In 59/70 invasive isolates, the lic2A gene was OFF, but it is also OFF in the majority of carriage isolates (16/17; no significant difference using a two-tailed Mann-Whitney U Test; Figure 2). Lic2A is a galactosyltransferase, and in tandem with LgtC, is responsible for the addition of a digalactoside Galα(1-4)βGal moiety (24, 27) onto the LOS. Lic2A activity is
the responsible for the addition of the first galactose onto a glucose, providing a substrate for LgtC to add the second galactose (Figure 1A).

We demonstrate that the gene encoding an O-acetyltransferase, oafA, is generally OFF in carriage isolates, but is ON in the majority of invasive NTHi isolates. The oafA gene is ON in 47/70 invasive NTHi isolates (67%), but only ON in 4/17 carriage isolates (23%; P-value = 0.011 using a two-tailed Mann-Whitney U Test; Figure 2). OafA adds an O-acetyl group to the heptose antigen of the inner core of the LOS (Figure 1A), and it has previously been reported that this O-acetylation, i.e., oafA ON, is required for resistance to complement-mediated killing by the host-immune system (28). The oafA gene is also uniformly present in invasive isolates, but is absent in 2/17 carriage isolates. The uniform presence of oafA in invasive isolates indicates that all NTHi that are invasive have the potential to switch oafA ON.

**Discussion**

Our investigation of a large collection of invasive NTHi isolates has allowed us to determine if particular LOS biosynthetic genes are present, and have altered expression in sterile niches in the human host. Whilst five out of seven of these biosynthetic genes (lic1A, lic3A, lic3B, lex2A, and lgtC) show no significant correlation with an ON or OFF expression state during invasive infection, we demonstrate that lic2A remains OFF in invasive isolates, and oafA ON is statistically over-represented in invasive isolates, when compared to carriage isolates.

Our observation that lic2A is OFF in most invasive isolates is intriguing as this finding appears contradictory to earlier results. Expression of lic2A was previously demonstrated to confer resistance to human serum (31), and modification of the NTHi LOS inner core with a galactose by Lic2A has been shown to shield the cells from *in vitro* neutrophil-mediated killing assays when lic1A is phase-varied OFF, with this modification associated with
invasive NTHi isolates (32). However, our findings demonstrate that lic2A is OFF in the majority (59/70) of invasive NTHi isolates. Further work is required to identify what factors initially cause Lic2A expression for resistance to serum (licA2 ON), but then appear to either select against its expression (licA2 OFF), or do not require its further expression, during invasive disease.

We previously demonstrated that oafA OFF is selected for during otitis media (19), whereas this work demonstrates oafA ON occurs during invasive disease. Previous work with oafA expression in NTHi has demonstrated that O-acetylation of the LOS by OafA is required for resistance to complement-mediated killing by human serum (28). The differences in selection for oafA expression between two host-niches (OFF in the middle ear/ON for invasion and serum resistance) demonstrates the rapid adaptability afforded by phase-variable genes: transition to occupying the middle ear appears to favour oafA OFF (19), whereas oafA ON occurs during invasive disease and is required for resistance to serum. Interestingly, loss of the related O-acetyltransferase OafA in the human enteric pathogen _Salmonella enterica_ serovar typhimurium, which acetylates the O-antigen of LPS (33), leads to modulation of the immune response, and may aid immune-evasion (34). Therefore, it appears that acetylation of outer surface oligosaccharides is a common evolutionary mechanism of bacterial pathogens to avoid the immune response, and perhaps leads to increased virulence.

Modification of NTHi LOS with other glycan moieties has been shown to be important during pathogenesis. For examples, NTHi strains isolated from blood show a decreased phosphorylcholine (PCho) content on their LOS relative to nasopharyngeal strains, which leads to decreased binding of antibodies and C-reactive protein (35), which aids survival in blood. However, this study did not investigate if the decreased PCho content of these invasive isolates was due to phase-variation of Lic1A, the glycosyltransferase responsible for this modification (Figure 1). We did not see any switching of lic1A in our survey (Figure 2),
which implies that the decreased PCho content of the LOS of invasive isolates (35) could be
due to a variety of factors that likely includes, but not absolutely dependent on, lic1A
switching OFF. Addition of a ketodeoxyoctanoate (KDO) residue as the terminal sugar of
LOS rather than N-acetylneuraminic acid (Neu5Ac; Figure 1) is present during NTHi biofilm
formation in vivo (36), meaning this modification may chronic infection with NTHi.

Previous studies examining the role of LOS phase-variation in NTHi pathobiology during
infection of human volunteers have investigated the ON-OFF status of LOS biosynthetic
genes (19, 37), and have shown selection for particular ON-OFF states: lex2A and lic1A were
shown to switch from OFF to ON during nasopharyngeal colonisation (37). This lic1A
finding corroborates the finding that shows decreased PCho in invasive NTHi isolates relative
to strains from the nasopharynx (35). Our findings that oafA ON is selected for during
invasive infection, and that the lic2A OFF expression state predominates in both carriage and
invasive NTHi strains, adds an extra level to the complexity of the factors that result in NTHi
transitioning from benign carriage to causing overt disease. Whilst we cannot determine if
particular LOS structures resulting from the ON-OFF status of these genes leads to invasion,
or is actually selected for as NTHi moves to particular host niches, i.e., becomes invasive, our
work has determined that particular LOS modifications are more prevalent during invasive
NTHi disease.

Expression and/or acquisition of particular factors was hypothesised to lead to the emergence
of a particularly virulent clone of the closely related organism, H. influenzae biogroup
aegyptius (38), responsible for the acute and fatal invasive infection Brazilian purpuric fever
(BPF) (39). H. aegyptius was previously well characterised as a pathogen causing purulent
conjunctivitis, but the changes in the organism that were responsible for transition from
causing conjunctivitis to causing severe invasive disease are uncharacterised. Nevertheless,
several virulence factors were identified (40), with acquisition of particular outer-membrane
proteins (41), secretion of extracellular proteins (42), expression of certain adhesins (43), and differences in LOS structure (38) all hypothesised to result in BPF, but none were ever conclusively shown to be absolutely required for virulence (38). Our demonstration that oafA ON is statistically associated with invasive isolates of NTHi could serve as an indicator for the invasive potential of NTHi strains, and is one of the first genes shown to be associated with invasive NTHi disease. However, not all invasive isolates in our collection expressed oafA, and it is highly likely that there as yet other uncharacterised factors associated with invasive NTHi infection.

In summary, our work has demonstrated a link between phase-variation of particular LOS biosynthetic genes (oafA ON and lic2A OFF) and invasive disease caused by NTHi. Understanding the expression of these proteins and the structure of LOS during NTHi infection is particularly important, as knowledge of the factors involved in invasive NTHi disease will allow the design of better treatments, allow more accurate diagnosis of infection, and aid in the design of an efficacious and broadly effective vaccine.

**Materials and Methods**

**Bacterial strains and growth conditions**

Invasive NTHi strains used for this study were isolated from sterile sites in patients suffering *H. influenzae* infections in South-East Queensland over a fifteen year period (29). Information on age, site of isolation, and geographical location were all collected, but information on any comorbidity was not (29). The seventy isolates used in this study were selected to represent a broad random sample of the strains present in this collection. NTHi isolates were grown on BHI (oxoid) supplemented (sBHI) with hemin (1%) and NAD (2ug/ml) at 37°C in an atmosphere containing 5% (v/v) CO₂. Isolates were previously confirmed as NTHi using commercially available sera (Phadebact Haemophilus Test; MKL...
Whole genome sequences of each of the seventy isolates were used to perform a BLAST search with NTHi OMP P2 and P6 gene sequences in order to provide additional confirmation (data not shown).

Nasal (carriage) control samples were taken from the ORChID collection, a prospective, birth cohort study of infants in South-East Queensland where daily symptoms were recorded and weekly nasal swabs were collected from 158 infants during their first two years of life (30). All samples used as carriage controls are from infants demonstrating no overt symptoms of respiratory illnesses either two weeks before or two weeks after sampling (44).

**DNA preparation, manipulation and analysis**

Bacterial genomic DNA from invasive isolates was prepared by boiling a 1 µl loop of each NTHi isolate in 200 µl TE buffer for 20 mins, removing the debris by centrifugation (14,000xg 5 mins) and collecting the supernatant, which contained genomic DNA. DNA from the ORChID carriage controls samples were isolated as described previously (45). 1 µl of each DNA preparation was used in each PCR reaction. PCR primers were purchased from Integrated DNA Technologies (IDT; Singapore). Primers are described in Table 1. Multiplex PCR was carried out in 25 µl reactions using GoTaq DNA polymerase (Promega) according to the manufacturer’s instructions. Cycle conditions were as follows: Initial Denaturation at 95°C for 2 mins, followed by 30 cycles of Denaturation at 95°C 30 secs, Anneal at 52°C for 30 secs, Extension at 72°C for 30 secs, with a final Extension at 72°C for 5 minutes. Samples were checked for multiplex products on 2% (w/v) agarose gels buffered with 1xTBE. DNA fragments were sized using the GeneScan system (Applied Biosystems International), at the Australian Genome Research facility (AGRF, Brisbane), and traces analysed using PeakScanner software (Applied Biosystems International). Where a PCR product could not be produced for a particular gene in an isolate, we analysed the genome sequence available for the invasive collection (PRJEB18702). An illustration of the fragment analysis PCR
methodology, and an example of a GeneScan trace and PeakScanner quantification are shown in Figure 1B and C, respectively. The results shown in Figure 2 indicate whether the genes investigated were ON (>70% ON; green), OFF (>70% OFF; red) or mixed ON and OFF (orange). This was determined from the number of nucleotide repeats in the SSR present in each gene (based on amplicon peak size), and calibrated using previous studies that have demonstrated the relationship between SSR length present in these seven LOS biosynthetic genes and gene expression status (19).

Acknowledgements

This work is supported by Australian Research Council (ARC) Discovery Project 180100976 to JMA, National Health and Medical Research Council (NHMRC; Australia) Project Grant 1099279 to KLS and JMA, Career Development Fellowship to KLS, and Program Grant 1071659 and Principal Research Fellowship 1138466 to MPJ. The ORChID collection was supported by NHMRC Project Grant GNT615700. ZNP is supported by a Griffith University Senior Deputy Vice Chancellor (SDVC) PhD Scholarship. The authors would like to acknowledge all pathology laboratories, private and public for referral of invasive isolates and Vicki Hicks, Kelly Progomet and Lawrence Ariotti of Public Health Microbiology, Forensic and Scientific Services, Queensland Department of Health for their work in the culture, identification and serotyping of referred invasive isolates.

Conflict of interest - none
References

1. Murphy TF, Faden H, Bakaletz LO, Kyd JM, Forsgren A, Campos J, Virji M, Pelton SI. 2009. Nontypeable Haemophilus influenzae as a pathogen in children. Pediatr Infect Dis J 28:43-8.

2. Sethi S, Murphy TF. 2008. Infection in the pathogenesis and course of chronic obstructive pulmonary disease. N Engl J Med 359:2355-65.

3. Van Eldere J, Slack MP, Ladhani S, Cripps AW. 2014. Non-typeable Haemophilus influenzae, an under-recognised pathogen. Lancet Infect Dis 14:1281-92.

4. Johnson RH. 1988. Community-acquired pneumonia: etiology, diagnosis, and treatment. Clin Ther 10:568-73.

5. Langereis JD, de Jonge MI. 2015. Invasive Disease Caused by Nontypeable Haemophilus influenzae. Emerging Infectious Diseases 21:1711-1718.

6. Agrawal A, Murphy TF. 2011. Haemophilus influenzae infections in the H. influenzae type b conjugate vaccine era. J Clin Microbiol 49:3728-32.

7. Gkentzi D, Slack MP, Ladhani SN. 2012. The burden of nonencapsulated Haemophilus influenzae in children and potential for prevention. Curr Opin Infect Dis 25:266-72.

8. Naito S, Takeuchi N, Ohkusu M, Takahashi-Nakaguchi A, Takahashi H, Imuta N, Nishi J, Shibayama K, Matsuoka M, Sasaki Y, Ishiwada N. 2018. Clinical and Bacteriologic Analysis of Nontypeable Haemophilus influenzae Strains Isolated from Children with Invasive Diseases in Japan from 2008 to 2015. J Clin Microbiol 56.

9. Ladhani S, Slack MP, Heath PT, von Gottberg A, Chandra M, Ramsay ME. 2010. Invasive Haemophilus influenzae Infection Disease, Europe, 1996-2006. Emerg Infect Dis 16:455-63.

10. Collins S, Vickers A, Ladhani SN, Flynn S, Platt S, Ramsay ME, Litt DJ, Slack MP. 2016. Clinical and Molecular Epidemiology of Childhood Invasive Nontypeable Haemophilus influenzae Disease in England and Wales. Pediatr Infect Dis J 35:e76-84.

11. Tribuddharat C, Sritruengfung S. 2017. Multiple drug resistance in Haemophilus influenzae isolated from patients in Bangkok, Thailand. J Glob Antimicrob Resist 9:121-123.

12. Atkinson CT, Kunde DA, Tristram SG. 2017. Expression of acquired macrolide resistance genes in Haemophilus influenzae. J Antimicrob Chemother 72:3298-3301.

13. Giufre M, Cardines R, Accogli M, Pardini M, Cerquetti M. 2013. Identification of Haemophilus influenzae Clones Associated with Invasive Disease a Decade after Introduction of H. influenzae Serotype b Vaccination in Italy. Clinical and Vaccine Immunology 20:1223-1229.

14. Satola SW, Napier B, Farley MM. 2008. Association of IS1016 with the hia adhesin gene and biotypes V and I in invasive nontypeable Haemophilus influenzae. Infect Immun 76:5221-7.

15. Moxon R, Bayliss C, Hood D. 2006. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. Ann Rev Genet 40:307-333.

16. Bayliss CD, Bidmos FA, Anjum A, Manchev VT, Richards RL, Grossier J, Wooldridge KG, Ketley JM, Barrow PA, Jones MA, Tretyakov MV. 2012. Phase variable genes of Campylobacter jejuni exhibit high mutation rates and specific mutational patterns but mutability is not the major determinant of population structure during host colonization. Nucleic Acids Res 40:5876-5889.

17. Cox EC. 1976. Bacterial mutator genes and the control of spontaneous mutation. Annu Rev Genet 10:135-56.

18. Farabaugh PJ, Schmeissner U, Hofer M, Miller JH. 1978. Genetic studies of the lac repressor. J Mol Biol 126:847-863.

19. Fox KL, Atack JM, Srikhanta YN, Eckert A, Novotny LA, Bakaletz LO, Jennings MP. 2014. Selection for phase variation of LOS biosynthetic genes frequently occurs in progression of non-typeable Haemophilus influenzae infection from the nasopharynx to the middle ear of human patients. PLoS One 9:e90505.

20. Weiser JN, Maskell DJ, Butler PD, Lindberg AA, Moxon ER. 1990. Characterization of repetitive sequences controlling phase variation of Haemophilus influenzae lipopolysaccharide. J Bacteriol 172:3304-9.
21. van Belkum A, Scherer S, van Leeuwen W, Willemse D, van Alphen L, Verbrugh H. 1997. Variable number of tandem repeats in clinical strains of Haemophilus influenzae. Infect Immun 65:5017-27.

22. van Belkum A, Scherer S, van Alphen L, Verbrugh H. 1998. Short-sequence DNA repeats in prokaryotic genomes. Microbiol Mol Biol Rev 62:275-293.

23. Weiser JN, Lindberg AA, Manning EJ, Hansen EJ, Moxon ER. 1989. Identification of a chromosomal locus for expression of lipopolysaccharide epitopes in Haemophilus influenzae. Infect Immun 57:3045-3052.

24. High NJ, Deadman ME, Moxon ER. 1993. The role of a repetitive DNA motif (5'-CAAT-3') in the variable expression of the Haemophilus influenzae lipopolysaccharide epitope alpha Gal(1-4)beta Gal. Mol Microbiol 9:1275-82.

25. Fox KL, Cox AD, Gilbert M, Wakarchuk WW, Li J, Makepeace K, Richards JC, Moxon ER, Hood DW. 2006. Identification of a Bifunctional Lipopolysaccharide Sialyltransferase in Haemophilus influenzae: INCORPORATION OF DISIALIC ACID. J Biol Chem 281:40024-40032.

26. Jarosik GP, Hansen EJ. 1994. Identification of a new locus involved in expression of Haemophilus influenzae type b lipooligosaccharide. Infect Immun 62:4861-4867.

27. Hood DW, Deadman ME, Jennings MP, Biseric M, Fleischmann RD, Venter JC, Moxon ER. 1996. DNA repeats identify novel virulence genes in Haemophilus influenzae. Proc Natl Acad Sci U S A 93:11121-5.

28. Fox KL, Yildirim HH, Deadman ME, Schweda EK, Moxon ER, Hood DW. 2005. Novel lipopolysaccharide biosynthetic genes containing tetranucleotide repeats in Haemophilus influenzae, identification of a gene for adding O-acetyl groups. Mol Microbiol 58:207-16.

29. Staples M, Graham RM, Jennison AV. 2017. Characterisation of invasive clinical Haemophilus influenzae isolates in Queensland, Australia using whole-genome sequencing. Epidemiol Infect doi:10.1017/s0950268817000430.

30. Lambert SB, Ware RS, Cook AL, Maguire FA, Whiteley DM, Bialasiewicz S, Mackay IM, Wang D, Sloots TP, Nissen MD, Grimwood K. 2012. Observational Research in Childhood Infectious Diseases (ORChID): a dynamic birth cohort study. BMJ Open 2: doi:10.1136/bmjopen-2012-002134.

31. Dixon K, Bayliss CD, Makepeace K, Moxon ER, Hood DW. 2007. Identification of the Functional Initiation Codons of a Phase-Variable Gene of Haemophilus influenzae, lic2A, with the Potential for Differential Expression. J Bacteriol 189:511-521.

32. Langereis JD, Weiser JN. 2014. Shielding of a lipooligosaccharide IgM epitope allows evasion of neutrophil-mediated killing of an invasive strain of nontypeable Haemophilus influenzae. MBio 5:e01478-14.

33. Slauch JM, Lee AA, Mahan MJ, Mekalanos JJ. 1996. Molecular characterization of the oafA locus responsible for acetylation of Salmonella typhimurium O-antigen: oafA is a member of a family of integral membrane trans-acylases. J Bacteriol 178:5904-9.

34. Slauch JM, Mahan MJ, Michetti P, Neutra MR, Mekalanos JJ. 1995. Acetylation (O-factor 5) affects the structural and immunological properties of Salmonella typhimurium lipopolysaccharide O antigen. Infect Immun 63:437-41.

35. Langereis JD, Cremers AJH, Vissers M, van Beek J, Meis JF, de Jonge ML. 2019. Nontypeable Haemophilus influenzae Invasive Blood Isolates Are Mainly Phosphorylcholine Negative and Show Decreased Complement-Mediated Killing That Is Associated with Lower Binding of IgM and CRP in Comparison to Colonizing Isolates from the Oropharynx. Infect Immun 87.

36. Apicella MA, Coffin J, Ketterer M, Post DMB, Day CJ, Jen FE, Jennings MP. 2018. Nontypeable Haemophilus influenzae Lipooligosaccharide Expresses a Terminal Ketodeoxyoctanoate In Vivo, Which Can Be Used as a Target for Bactericidal Antibody. MBio 9.

37. Poole J, Foster E, Chaloner K, Hunt J, Jennings MP, Bair T, Knudtson K, Christensen E, Munson RS, Jr., Winokur PL, Apicella MA. 2013. Analysis of nontypeable Haemophilus influenzae phase variable genes during experimental human nasopharyngeal colonization. J Infect Dis 208:720-727.
38. Harrison LH, Simonsen V, Waldman EA. 2008. Emergence and disappearance of a virulent clone of Haemophilus influenzae biogroup aegyptius, cause of Brazilian purpuric fever. Clin Microbiol Rev 21:594-605.

39. Harrison LH, da Silva GA, Pittman M, Fleming DW, Vranjac A, Broome CV. 1989. Epidemiology and clinical spectrum of Brazilian purpuric fever. Brazilian Purpuric Fever Study Group. J Clin Microbiol 27:599-604.

40. Carlone GM, Gorelkin L, Gheesling LL, Hoiseth SK, Mulks MH, O'Connor SP, Weyant RS, Myrick JE, Mayer LW, Arko RJ. 1989. Potential virulence factors of Haemophilus influenzae biogroup aegyptius in Brazilian purpuric fever. The Brazilian Purpuric Fever Study Group. Pediatr Infect Dis J 8:245-7.

41. Li MS, Farrant JL, Langford PR, Kroll JS. 2003. Identification and characterization of genomic loci unique to the Brazilian purpuric fever clonal group of H. influenzae biogroup aegyptius: functionality explored using meningococcal homology. Mol Microbiol 47:1101-11.

42. Barbosa SF, Hoshino-Shimizu S, Alkmin M, Goto H. 2003. Implications of Haemophilus influenzae biogroup aegyptius hemagglutinins in the pathogenesis of Brazilian purpuric fever. J Infect Dis 188:74-80.

43. Strouts FR, Power P, Croucher NJ, Corton N, van Tonder A, Quail MA, Langford PR, Hudson MJ, Parkhill J, Kroll JS, Bentley SD. 2012. Lineage-specific virulence determinants of Haemophilus influenzae biogroup aegyptius. Emerg Infect Dis 18:449-57.

44. Sarna M, Ware RS, Sloots TP, Nissen MD, Grimwood K, Lambert SB. 2016. The burden of community-managed acute respiratory infections in the first 2-years of life. Pediatric Pulmonology 51:1336-1346.

45. Sarna M, Lambert SB, Sloots TP, Whiley DM, Alsaleh A, Mhango L, Bialasiewicz S, Wang D, Nissen MD, Grimwood K, Ware RS. 2018. Viruses causing lower respiratory symptoms in young children: findings from the ORChID birth cohort. Thorax 73:969-979.
Figure 1 - Illustration of NTHi LOS structure and fragment analysis methodology. A) Schematic representation of NTHi LOS and the roles of the glycosyltransferases encoded by the seven phase-variable loci studied in this work: Lic1A = phosphorylcholine transferase, Lic2A = galactosyltransferase, Lic3A and Lic3B = sialyltransferases, Lex2A = glucosyltransferase, LgtC = galactosyltransferase, OafA = O-acetyltransferase (28). NTHi LOS contains 2-keto-3-deoxyoctulosonic acid (KDO); pyrophosphoethanolamine (PPEtn); phosphoethanolamine (PEtn); heptose (Hep); galactose (Gal); glucose (Glc); phosphocholine (PCho); Neu5Ac (N-acetylneuraminic acid); and O-acetyl group (OAc). LOS structure is therefore dependent on the ON/OFF status of each of these seven genes; B) An illustration of the PCR technique used to survey the repeat tract length of a phase-variable gene, in this case oafA, which contains a variable length SSR tract made up of a GCAAn repeat (green box). Primers are designed to bind either side of this repeat tract, with the length of PCR product dependent on the number of GCAAn repeats present. Therefore a population will contain a mixture of different sized PCR products as the length of the repeat tract varies between individual bacterial cells. Fragments are then separated, sized, and the amount of each size quantified using an ABI GeneScan system by using a fluorescently labelled forward primer (green star); C) An example of a GeneScan fragment analysis trace, with the area under each peak representing the proportion of that fragment size (in bp) in the population. As we know what tract lengths lead to the ON or OFF status of each gene, we can then determine the proportion of the population that is ON or OFF based on this quantification.

Figure 2 - heat-map showing the expression status of each of the seven phase-variable LOS biosynthetic loci assessed in this study. Seventy invasive NTHi isolates (29) and seventeen NTHi carriage isolates (30) were assessed for ON/OFF status using multiplexed fluorescent PCR. Fragment lengths were quantified using an ABI Genescan system, and
quantified using PeakScanner software. ON/OFF status was calculated as described previously (19). Green = >70% ON; Red = >70% OFF; Orange = mixed ON/OFF; blue = no repeat tract; grey = no gene (no product from multiple PCR attempts). All % ON and OFF values for each collection can be found in Supplementary Data 2.

Table 1 - Primers used in this study

| Gene | Repeat unit | Reps ON or OFF example | Primers | refs |
|------|-------------|------------------------|---------|------|
| lgtC | GACA        | 10 = ON; 11 & 12 = OFF | For: 5’-VIC-TCATCGAGCAAGGCATTG-3’ Rev: 5’-CTTACAGCTAATAAGGTGC-3’ | Fox et al, 2014 |
| lex2A | GCAA        | 10 & 11 = OFF; 12 = ON | For: 5’-NED-CGGAATTATGTATACAC-3’ Rev: 5’-GTTCCTTGTGTATAC-3’ | Fox et al, 2014 |
| lic2A | CAAT        | 10 = ON; 11 & 12 = OFF | For: 5’-FAM-ACTGAACGTGCAAA-3’ Rev: 5’-GCTAATTAACAGCCT-3’ | High et al, 1993 |
| lic1A | CAAT        | 10 & 11 = OFF; 12 = ON | For: 5’-VIC-CAAAAAATAACTTTAAGCGT-3’ Rev: 5’-AATGCTGATGAAGAAAATG-3’ | Fox et al, 2014 |
| lic3A | CAAT        | 10 & 11 = OFF; 12 = ON | For: 5’-NED-ATTACCTGCAATAATGACAG-3’ Rev: 5’-TATTCAATGAACGTTAGAAT-3’ | Van Belkum et al, 1997; this study |
| gene | specific primer sequences | For: | Rev: | Reference |
|------|--------------------------|------|------|-----------|
| Lic3A specific | 5’GCCAGTAGTCGCAAAAGTGTC-3’ |  |  | Van Belkum *et al.*, 1997; this study |
| lic3B | CAAT 11 = ON; 12 & 13 = OFF |  |  | Fox *et al.*, 2005 |
| oafA | GCAA 9 & 10 = OFF; 11 = ON |  |  | Srikhanta *et al.*, 2005 |
| modA | AGCC or AGTC 10 = ON; 11 & 12 = OFF |  |  |  |

**Supplementary Data 1** - Repeat tract length and % ON-OFF for each gene in each strain

**Supplementary Data 2** - % ON-OFF-mixed ratios for each gene in each collection
| Sample | Alg1C | HN7F | cas/C | Alc1A | Alc2A | Alc3A | Alc4B | Source |
|--------|-------|------|-------|-------|-------|-------|-------|--------|
| H11    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H12    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H13    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H14    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Lung Tissue |
| H15    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H16    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H17    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H18    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H19    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H20    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H21    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H22    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H23    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H24    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H25    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H26    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H27    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H28    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H29    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H30    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H31    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H32    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H33    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H34    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H35    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H36    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H37    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H38    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H39    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H40    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H41    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H42    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H43    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H44    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H45    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H46    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H47    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H48    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H49    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H50    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H51    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H52    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H53    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H54    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H55    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H56    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H57    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H58    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H59    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H60    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H61    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H62    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H63    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |
| H64    | ON    | ON   | OFF   | OFF   | ON    | OFF   | OFF   | Blood  |

| Genus | TRPV5 | TRPV6 | TRPV7 | TRPV8 | TRPV9 |
|-------|-------|-------|-------|-------|-------|
| SW1152 | ON    | ON    | OFF   | OFF   | OFF   |
| SW1153 | ON    | ON    | OFF   | OFF   | OFF   |
| SW1154 | ON    | ON    | OFF   | OFF   | OFF   |
| SW1158 | ON    | ON    | OFF   | OFF   | OFF   |
| SW1160 | ON    | ON    | OFF   | OFF   | OFF   |
| SW1162 | ON    | ON    | OFF   | OFF   | OFF   |
| SW1174 | ON    | ON    | OFF   | OFF   | OFF   |
| SW1178 | ON    | ON    | OFF   | OFF   | OFF   |
| SW1185 | ON    | ON    | OFF   | OFF   | OFF   |
| SW1187 | ON    | ON    | OFF   | OFF   | OFF   |
| SW1202 | ON    | ON    | OFF   | OFF   | OFF   |
| SW1203 | ON    | ON    | OFF   | OFF   | OFF   |
| SW1205 | ON    | ON    | OFF   | OFF   | OFF   |
| SW1206 | ON    | ON    | OFF   | OFF   | OFF   |
| SW1207 | ON    | ON    | OFF   | OFF   | OFF   |
| SW1208 | ON    | ON    | OFF   | OFF   | OFF   |

**P-value**: 0.044 0.937 0.001 0.032 0.048 0.052 0.08

**Legend**:
- **ON**: mixed
- **OFF**: no gene
- **red**: no repeat tract (not phase-variable)