Published in final edited form as:

Infect Genet Evol. 2014 December ; 28: 223–232. doi:10.1016/j.meegid.2014.09.035.

Prevalence, distribution, and sequence diversity of *hmwA* among commensal and otitis media non-typeable *Haemophilus influenzae*

Gregg S. Davis^{a,1}, May Patel^b, James Hammond^a, Lixin Zhang^a, Suzanne Dawid^{b,c}, Carl F. Marrs^a, and Janet R. Gilsdorf^{a,b}

Gregg S. Davis: gsdavis@gwu.edu; May Patel: mpatel@umich.edu; James Hammond: jamhenha@umich.edu; Lixin Zhang: lxzhang@umich.edu; Suzanne Dawid: sdawid@umich.edu; Carl F. Marrs: cfmarrs@umich.edu; Janet R. Gilsdorf: gilsdorf@med.umich.edu

^aDepartment of Epidemiology, University of Michigan School of Public Health, 1415 Washington Heights, Ann Arbor, MI 48109, USA

^bDepartment of Pediatrics, University of Michigan Medical School, D5101 Medical Professional Building, Ann Arbor, MI 48109-5718, USA

^cDepartment of Microbiology and Immunology, University of Michigan Medical School, 1150 W. Medical Center Drive, Ann Arbor, MI 48109, USA

Abstract

Nontypeable *Haemophilus influenzae* (NTHi) are Gram-negative coccobacilli that colonize the human pharynx, their only known natural reservoir. Adherence to the host epithelium facilitates NTHi colonization and marks one of the first steps in NTHi pathogenesis. Epithelial cell attachment is mediated, in part, by a pair of high molecular weight (HMW) adhesins that are highly immunogenic, antigenically diverse, and display a wide range of amino acid diversity both within and between isolates. In this study, the prevalence of *hmwA*, which encodes the HMW adhesin, was determined for a collection of 170 NTHi isolates recovered from the middle ears of children with otitis media (OM isolates) or throats of healthy children (commensal isolates) from Finland, Israel, and the U. S. Overall, *hmwA* was detected in 61% of NTHi isolates and was significantly more prevalent ($P = 0.004$) among OM isolates than among commensal isolates; the prevalence ratio comparing *hmwA* prevalence among ear isolates with that of commensal isolates was 1.47 (95% CI (1.12, 1.92)). Ninety-five percent (98/103) of the *hmwA*-positive NTHi isolates possessed two *hmw* loci. To advance our understanding of *hmwA* binding sequence diversity, we determined the DNA sequence of the *hmwA* binding region of 33 isolates from this collection. The average amino acid identity across all *hmwA* sequences was 62%. Phylogenetic analyses of the *hmwA* binding revealed four distinct sequence clusters, and the majority of *hmwA* sequences

© 2014 Elsevier B.V. All rights reserved.

¹Current Address: Department of Environmental and Occupational Health, Milken Institute School of Public Health, The George Washington University, 950 New Hampshire Ave., NW, 4th Floor, Washington, DC 20052

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

All authors report no conflicts of interest.

(83%) belonged to one of two dominant sequence clusters. *hmwA* sequences did not cluster by chromosomal location, geographic region, or disease status.

Keywords

high molecular weight adhesins; evolution; phylogenetics; adherence; NTHi; acute otitis media

1. Introduction

Nontypeable *Haemophilus influenzae* (NTHi) are Gram-negative bacteria that colonize the human pharynx, their only known natural reservoir. Although generally considered a commensal, NTHi are capable of causing localized infections of the upper and lower respiratory tracts, such as acute otitis media, sinusitis, and bronchitis, as well as, more rarely, severe invasive infections such as pneumonia, bacteremia, and meningitis. NTHi colonization is initiated by adherence of bacterial cells to the host epithelium and, since disease-causing isolates are thought to arise from the isolates colonizing the pharynx, adherence also marks the first step in NTHi pathogenesis. NTHi produce a wide array of surface exposed adhesins that mediate attachment to epithelial cells (St Geme and Cutter, 1995, 1996; St Geme et al., 1994; St Geme et al., 1993; Swords et al., 2000). High molecular weight (HMW) adhesins, which have been reported in 38 to 80% of NTHi isolates (Barenkamp, 1992; Barenkamp and Leininger, 1992; Ecevit et al., 2004; Erwin et al., 2005; Erwin et al., 2008; St Geme et al., 1998; van Schilfgaarde et al., 2000), are among the predominant non-pilin NTHi adhesins and are widely distributed across NTHi phylogenetic groups as defined by multi-locus sequence typing (MLST) (Erwin et al., 2008) and multi-locus enzyme electrophoresis (MLEE) (St Geme et al., 1998).

The HMW adhesins are a family of paralogous proteins encoded by the *hmw* locus, which is present in two complete copies designated *hmw1* and *hmw2*, located discontinuously on the NTHi chromosome (Buscher et al., 2004). Each *hmw* locus encodes three proteins, *hmwA*, *hmwB*, and *hmwC* and both the gene content and the chromosomal locations of the *hmw* loci are conserved across isolates (Barenkamp and St Geme, 1994; Buscher et al., 2004). *hmwA* encodes the functional HMW adhesins while *hmwB* and *hmwC* encode proteins required for maturation, glycosylation, and secretion of mature HMW adhesins (Barenkamp and Leininger, 1992; Barenkamp and St Geme, 1994; Grass et al., 2003; Grass and St Geme, 2000; St Geme and Grass, 1998). The promoter regions upstream of each *hmwA* possess a variable number of heptanucleotide simple sequence repeats (SSRs) that mediate phase variation (Dawid et al., 1999)—as the number of repeats increases, *hmwA* transcription decreases. The *hmw1* and *hmw2* loci of NTHi strain 12 are located downstream of HI1679 and HI1598, respectively, as specified with respect to *H. influenzae* strain Rd (Barenkamp and Leininger, 1992).

HMW adhesins display a wide range of amino acid diversity both within and between isolates (Barenkamp and Leininger, 1992; Buscher et al., 2004; Dawid et al., 2001; Giufre et al., 2008). Rather than being evenly distributed throughout each mature HMW adhesin, the amino acid diversity is localized to the binding domain, the region of the adhesin that

interacts with the host epithelium (Barenkamp and Leininger, 1992; Dawid et al., 2001; Ecevit et al., 2005; Giufre et al., 2006). Amino acid diversity within the HMW binding domain likely serves two functions. First, as evidenced by characterization of five distinct NTHi *in vitro* adherence patterns based on the inhibiting effect of dextran sulfate and two different cell lines, amino acid diversity helps define NTHi tissue tropism (van Schilfgaarde et al., 2000). Most isolates analyzed to date, however, display adherence patterns similar to that of the prototypic NTHi strain 12 in which one adhesin confers HMW1-like adherence, characterized by high level adherence to Chang cells, Hep-2, HaCaT, and NCI-H292 cells mediated by interactions with α -2,3 N-linked sialic acids, and the other HMW2-like adherence, characterized by adherence to only HaCaT and NCI-H292 cells via a 2,3 sialic acid independent mechanism, the exact nature of which has not been determined (Buscher et al., 2004; St Geme et al., 1993; van Schilfgaarde et al., 2000). Secondly, because the HMW adhesins are exposed to the external environment and directly interact with host tissues, amino acid diversity likely plays a role in immune evasion.

All current, publicly available, *hmwA* DNA sequence data are generated exclusively from disease isolates (Barenkamp and Leininger, 1992; Buscher et al., 2004; Ecevit et al., 2005; Giufre et al., 2006; Hogg et al., 2007; van Schilfgaarde et al., 2000). Disease isolates, however, represent only a subset of the total NTHi population, the majority of which consists of commensal isolates residing within the pharynges of healthy individuals. In this study, we utilized a well characterized collection of NTHi isolates isolated from the middle ears of children with otitis media (OM isolates) or from the throats of healthy children (commensal isolates) from Finland, Israel, and the U. S. (Lacross et al., 2013). We determined that the overall *hmwA* prevalence, determined by a polymerase chain reaction (PCR), in this collection of isolates was 54% and that OM isolates were significantly more likely to possess *hmwA* than were commensal isolates. We then sequenced the binding regions of the two copies of *hmwA* from 33 isolates widely distributed across a multilocus sequence typing (MLST) phylogeny estimated from this isolate collection (Lacross et al., 2013). *hmwA* sequence analysis revealed a wide range of genetic diversity both within and between NTHi isolates and phylogenetic analyses suggested that *hmwA* binding region form four distinct sequence clusters.

2. Materials and Methods

2.1. Strain collection

The NTHi isolates used in this study have been described in detail elsewhere (Lacross et al., 2013). In brief, the collection contains 170 NTHi isolates isolated between 1994 – 2002 from children seven years of age or under living in Finland, Israel, or the United States (for details, see Table 1 of Lacross, et al. (Lacross et al., 2013)). Ninetyfive of the isolates were collected from the middle ears of children with acute otitis media (“OM” isolates) and the remaining 75 were collected from the throats or nasopharynges of otherwise healthy children (“commensal” isolates) (Table S1).

2.2. *hmwA* nomenclature

The HMW adhesins were first characterized in NTHi strain 12, which is considered the prototypic HMW adhesin-positive strain (Barenkamp and Bodor, 1990; Barenkamp and Leininger, 1992; Barenkamp and St Geme, 1994). The two HMW adhesins produced by strain 12 differed in their molecular weights as determined by polyacrylamide gel electrophoresis; HMW1 (encoded by *hmwA*₁₆₇₉) migrated at approximately 125 kDa and HMW2 (encoded by *hmwA*₁₅₉₈) migrated at approximately 120 kDa. This difference in molecular weight was the basis of the original naming system in which the genes encoding the 125 kDa and 120 kDa adhesins were named *hmw1A* and *hmw2A*, respectively (Barenkamp and Leininger, 1992). In a later study, Buscher *et al.* (Buscher et al., 2004), determined the exact chromosomal locations of the *hmw1* and *hmw2* loci in NTHi strain 12 and defined each locus with reference to the open reading frame (ORF) located immediately upstream of each *hmw* locus in the fully sequenced *H. influenzae* strain Rd (Fleischmann et al., 1995). The NTHi strain 12 *hmw1* locus resides adjacent to Rd ORF HI1679, contains *hmw1A*, *B*, and *C*, and encodes the HMW1 adhesin, the strain 12 *hmw2* locus resides adjacent to Rd ORF HI1598, contains *hmw2A*, *B*, and *C*, and encodes the HMW2 adhesin (Buscher et al., 2004).

In this study, we assign each *hmwA* a locus-specific designation, *hmwA*₁₅₉₈ or *hmwA*₁₆₇₉, which corresponds to its chromosomal location relative to *H. influenzae* strain Rd ORF HI1598 or Rd ORF HI1679, respectively. Since we did not determine HMW adherence profiles in this study, designating *hmwA* by chromosomal location allows us to track each *hmwA* locus without reference to the respective proteins' adherence characteristics. With reference to NTHi strain 12, our *hmwA*₁₆₇₉ and *hmwA*₁₅₉₈ designations correspond to strain 12 *hmw1A* and *hmw2A*, respectively.

2.3. *hmwA*-leader PCR screen

To determine *hmwA* prevalence, DNA from whole cell lysates of all 170 NTHi isolates were interrogated with a single pair of PCR primers (Table 1) that annealed to a conserved region of both *hmwA*₁₆₇₉ and *hmwA*₁₅₉₈ signal sequences (Fig. 1). Whole cell lysate was prepared by suspending several bacterial colonies, collected with a sterile cotton swab, from a fresh plate into Tris-EDTA (TE, [pH 7.5]), heating for 10 min at 100 °C in a thermocycler, centrifuging for 10 min, and removing the supernatant, which was stored at – 20 °C until used for PCR. Conserved signal sequence regions were identified from an alignment of all 13 publicly available full-length *hmwA* sequences (Table S2). For routine *hmwA*-leader screening, PCR was performed with *Taq* polymerase (New England Biolabs, Ipswich, MA, cat. no. M0267). Each 20 µl PCR reaction consisted of 1× ThermoPol Reaction buffer, 0.2 mM of each dNTP, 0.2 µM of each primer, 2.0 units of *Taq* polymerase, and 2 µl crude lysate (template). Reactions were subjected to an initial denaturation step of 2 minutes at 95 °C followed by 30 amplification cycles of 95 °C for 30 seconds, 58 °C for 30 seconds and 72 °C for 1 minute. Amplicons were separated on 1.5% agarose gels in 1× Tris-Acetate-EDTA (TAE) buffer and visualized with ethidium bromide staining. NTHi strain 12 and strain 11 served as *hmwA*-positive and –negative controls, respectively.

2.4. Library on a slide microarray studies

H. influenzae library-on-a-slide (LOS) microarrays, consisting of *H. influenzae* genomic DNA arrayed on a glass slide, were prepared, tested, and analyzed as previously described (Sandstedt et al., 2008; Zhang et al., 2004). In brief, total genomic DNA from each *H. influenzae* strain was isolated and spotted in duplicate onto membrane coated slides. First, a digoxigenin labeled genomic DNA quantity control probe consisting of equal amounts of each of the seven *H. influenzae* multilocus sequence typing (MLST) genes (<http://Haemophilus.mlst.net>) and *pepN* was hybridized to the microarray (Sandstedt et al., 2008). The intensity of control probe hybridization to each spot was determined using Spotfinder v. 3.1.1 (<http://www.tm4.org/spotfinder.html>) and duplicate spots were averaged using MIDAS v.2.19 (<http://www.tm4.org/midas.html>) (Saeed et al., 2003). The slides were stripped and then rehybridized with a *hmwA*-leader region fluorescein labeled probe (Kong et al., 2006). The *hmwA* probe was “mixed probe” that contained equal concentrations of either *hmwA* amplified from multiple isolates. The probe was PCR amplified from genomic DNA with primers *hmwA_leader_F* and *hmwA_leader_R* (Fig. 1 and Table 1). Signal intensities were calculated as described above and the ratio of the log transformed *hmwA* signal to the MLST-*pepN* concentration control signal was calculated for each spot and the distribution of intensities was modeled using programs written in the statistical software “R” (<http://www.r-project.org>; (R Core Team, 2014)); spots were classified as positive, negative, or uncertain as described in Sandstedt, *et al.* (Sandstedt et al., 2008).

2.5. *hmwA* Southern blot analysis

A subset of *hmwA*-leader positive isolates (Table S1) were subjected to Southern blotting for detection of *hmwA* as follows: five micrograms of NTHi genomic DNA was digested with either *Bgl*III or *Eco* RI, separated on a 0.8% agarose gel in 1× TAE, and transferred by capillary blotting onto Hybond-H+ (Amersham Biosciences, Piscataway, NJ). NTHi isolates 12 and 11 served as *hmwA*-positive and –negative controls, respectively. Following UV-crosslinking, genomic DNA was interrogated with a *hmwA* leader probe (Fig. 1) prepared as described in section 2.4. Briefly, the blot was incubated in 25 mls of 1× PerfectHyb Plus Hybridization buffer (Sigma-Aldrich, St. Louis, MO) for three hours at 68°C prior to overnight hybridization with the *hmwA* probe at 68°C in PerfectHyb Plus. Following hybridization, the membrane was subjected to two five-minute washes in low stringency buffer (2× SSC, 0.1% SDS) and two 20-minute washes in high stringency buffer (0.2× SSC, 0.1% SDS); both washes were performed at 68°C. The blot was briefly rinsed with buffer (0.1M Tris [pH 7.5], 0.3 M NaCl) and then incubated for one hour at room temperature in 20 mls of block solution (0.1M Tris buffer [pH 7.5], 0.3M NaCl, 10% liquid blocking reagent (Amersham Biosciences, Piscataway, NJ)). The hybridized *hmwA*-leader probe was detected by incubating the blot for one hour at room temperature in a 1:5000 dilution of Anti-Fluorescein-AP Fab fragments (Roche Diagnostics, Mannheim, Germany) in block solution. Prior to color development with NBT/BCIP (Roche Diagnostics, Mannheim, Germany), the blot was subjected to three washes, ten minutes each, in 100 mls of wash solution (0.1M Tris (pH 7.5), 0.3 M NaCl, 0.3% Tween 20) followed by a final five minute incubation with high pH buffer (0.1M Tris (pH 9.5), 0.1M NaCl, 0.01 M MgCl₂). *hmwA*-leader probe was

visualized following color development with NBT/BCIP (Roche Diagnostics, Mannheim, Germany).

2.6. *hmwA*₁₆₇₉ and *hmwA*₁₅₉₈ locus-specific PCR amplification

Total genomic DNA was isolated from all *hmwA*-leader positive NTHi isolates and subjected to further analysis with *hmwA*₁₆₇₉ and *hmwA*₁₅₉₈ specific PCR primer pairs (Table 1). Genomic DNA was isolated using the Wizard genomic DNA purification kit (Promega, Madison, WI., cat. no. A1120) according to the manufacturer's instructions, resuspended in 1× Tris-EDTA buffer (10 mM Tris-HCL and 1mM EDTA at pH 8), and stored at 4 °C until needed. For *hmwA*-specific reactions, PCR was performed with Phusion High-Fidelity DNA polymerase (New England Biolabs, Ipswich, MA, cat. no. M0530). Each 25 µl PCR assay consisted of 1× Phusion HF buffer, 125 ng of genomic DNA template, 0.2 mM of each dNTP, and 0.2 µM of each primer, and 0.5 units of Phusion polymerase. Reactions were subjected to an initial denaturation step of 2 minutes at 98 °C followed by 30 amplification cycles of 98 °C for 15 seconds, 65 °C for 15 seconds and 72 °C for 3 minutes. Amplicons were separated on 1.0% agarose gels in 1× Tris-Acetate-EDTA (TAE) buffer and visualized under UV illumination following ethidium bromide staining. In the majority of isolates tested, primers A1FL5.1 and *hmwA* core_1R successfully amplified *hmwA*₁₆₇₉ and A2FL5.2 and *hmwA* core_1R successfully amplified *hmwA*₁₅₉₈. When one or both of the locus-specific PCRs failed to detect a *hmw* locus, additional primer pairs were tested. The primer pairs *hmwIA* F_ORF5-1 and *hmw2A* F_ORF5-1 coupled with *hmwA* R_HMWB3R were the most likely to detect *hmw* loci (Table 1). In spite of repeated attempts to optimize the PCR reactions, and the design of alternative primer pairs (not shown), we were unable to detect one, or both, loci in some isolates. NTHi strain 12 and strain 11 served as *hmwA* positive and –negative controls, respectively.

2.7. *hmwA* binding region amplification, cloning, and sequencing

The *hmwA*₁₆₇₉ and *hmwA*₁₅₉₈ binding regions were sequenced from a subset of the NTHi isolates used in this study (Fig. 2 and Table S2). By selecting sequences widely distributed across the phylogeny it was our primary intention to capture a wide range of *hmwA* diversity. First, nearly full-length *hmwA*₁₆₇₉ and *hmwA*₁₅₉₈ was specifically amplified from each selected NTHi strain using PCR primers *hmwA* 1F_ORF5-1 or *hmwA* 2F_ORF5-1, respectively, along with the reverse primer *hmwA*_core_1R which annealed to both loci (Table 1). Each full-length primary *hmwA* amplicon was then purified with the QIAquick PCR purification kit (QIAGEN, Valencia, CA, cat. no. 28104) according to the manufacturer's protocol, and *hmwA* binding regions were PCR amplified using Phusion High-Fidelity DNA polymerase and primers *hmwA* seqF_*SpeI* and *hmwA* seqR_*NotI* (Table 1), which annealed to shared conserved regions of *hmwA*₁₆₇₉ and *hmwA*₁₅₉₈. Each 25 µl PCR assay consisted of 1× Phusion HF buffer, 10 ng of purified primary *hmwA* PCR product, 0.2 mM of each dNTP, 0.2 µM of each primer, and 0.5 units of Phusion polymerase. Reactions were subjected to an initial denaturation step of 1 minute at 98 °C followed by 30 amplification cycles of 98 °C for 15 seconds, 65 °C for 15 seconds and 72 °C for 1.25 minutes. Amplicons were separated on 1.5% agarose gels in 1× Tris-Acetate-EDTA (TAE) buffer, stained with ethidium bromide, and visualized by UV transillumination.

Each *hwmA* binding region was cloned into Invitrogen's pCR4 sequencing vector with the Zero Blunt TOPO PCR cloning kit (Invitrogen, Carlsbad, CA, cat. no. K2875) in accordance with the manufacturer's protocols. Plasmid was isolated from overnight bacterial cultures with a GenCatch Plus plasmid DNA miniprep kit (Epoch Life Sciences, Sugar Land, TX; cat. no. 21-60050) following manufacturer's protocols. *hwmA* insert size was verified by restriction enzyme digestion followed by separation on a 1.5% agarose gel in 1× Tris-Acetate-EDTA (TAE) buffer and visualized with UV transillumination following ethidium bromide staining.

Sanger DNA sequencing was performed at the University of Michigan DNA Sequencing Core. Multiple sequencing reactions were required per *hwmA* core binding region and additional sequencing primers were synthesized as needed. PCR primers were designed using IDT PrimerQuest web interface (www.idtdna.com/Primerquest/Home/Index), sequence editing and contig assembly was performed with SeqMan Pro v8 (DNASTAR, Madison, WI). Thirteen publicly available NTHi *hwmA* nucleotide sequences, obtained from GenBank, were also included in the analyses (Table S2).

2.8. *hwmA* sequence analysis

hwmA nucleotide sequences were translated to amino acids, aligned with the computer program PRANKSTER (Loytynoja and Goldman), and then converted back to their original nucleotide sequences for analysis with the computer program MEGA5 (Tamura et al., 2011). Gaps were deleted from nucleotide alignments prior to analyses. Pairwise amino acid p-distances – the proportion of amino acids that vary between two sequences – were calculated for all sequence pairs using the pairwise gap deletion option. Maximum composite likelihood distances (Yang, 1994) were estimated from the *hwmA* nucleotide sequences. Rate heterogeneity between sites was modeled with a discrete gamma distribution with five rate categories; the gamma parameter value was estimated for each *hwmA* alignment under a general time reversible substitution model (Nei and Kumar, 2000). Standard errors of the distance measures, both amino acid and nucleotide, were estimated using a bootstrap method with 1000 replicates (Felsenstein, 1985).

Sequence alignments were tested for evidence of recombination with the PHI test (Buen et al., 2006), implemented in SplitsTree4 (Huson and Bryant, 2006). To account for the effects of recombination, relationships between *hwmA* sequences were assessed using the Neighbor-Net (Bryant and Moulton, 2004) method implemented in SplitsTree4 (Huson and Bryant, 2006). *hwmA* maximum likelihood phylogenies, implemented in MEGA5 (Tamura et al., 2011), were also estimated using a general time reversible model of sequence evolution; rate heterogeneity among sites was modeled with a five category discrete gamma distribution.

2.9. Statistical analyses

Relative comparisons between groups are reported as prevalence ratios (PR) with 95% confidence intervals (CI); commensal isolates served as the reference group for all comparisons. Prevalence ratios and 95% CI were estimated with the epiR software package implemented in R version 3.0.1 (R Core Team, 2014). Two tailed Fisher's exact tests,

implemented in R version 3.0.1 (R Core Team, 2014), were used to assess pairwise comparisons between groups. The threshold for statistical significance was $\alpha = 0.05$.

3. Results

3.1. *hmwA* prevalence

Initially, all 170 isolates in the collection (Lacross et al., 2013) (Table S1) were screened for the presence of *hmwA* by PCR with *hmwA*_leader PCR primers (Table 1). To account for the possibility of false-negative PCR results due to sequence diversity, all 170 isolates were also subjected to library-on-a-slide (LOS) hybridization with a *hmwA* leader-region probe (Fig. 1). Both screening methods targeted conserved regions of the *hmwA*₁₅₉₈ and *hmwA*₁₆₇₉ signal sequences and did not differentiate between the two *hmw* loci within an isolate. Across all isolates, the concordance between PCR and LOS methods was 86%, within each region concordance varied between 78% for isolates from Israel to 91% for U. S. isolates (Table S3). The number of discordant results was similar for OM ($n = 11$) and commensal isolates ($n = 13$). In total, 13 isolates tested PCR-negative/ LOS-positive, suggesting that variability within the PCR primer annealing region precluded *hmwA* amplification. All 13 PCR-negative/LOS-positive isolates tested positive by Southern blotting. Surprisingly, there were 11 PCR-positive/LOS-negative isolates, a result that is consistent with nucleotide variability across the length of the signal sequence but not within the primer annealing regions. Of these 11 isolates, nine encoded two *hmwA* loci as determined by locus-specific PCR, whereas, following Southern blotting, two were determined to be false-positive PCR results. Finally, three isolates, two of which were PCR-negative, failed to meet our quality control criteria during LOS analysis and were thus analyzed by Southern blot. One of the PCR-negative isolates hybridized with the *hwmA* probe, indicating a false-negative PCR, whereas PCR and Southern results were concordant for the two remaining isolates.

Overall, 103 (61%) isolates were *hmwA*-leader positive (Tables 2 and S1). *hmwA* prevalence, varied across geographic regions, ranging from 46% in Israel to 78% in Finland (Table 2). Isolates from Finland were significantly more likely to possess *hmwA* than were isolates from the U. S. ($P = 0.019$) or Israel ($P = 0.001$). There was, however, no significant difference ($P = 0.264$) in *hmwA* prevalence between isolates from the U. S. and Israel.

Taken together, OM isolates were significantly ($P = 0.004$) more likely to possess *hmwA* than were commensal isolates. The prevalence ratio (PR) comparing *hmwA* prevalence among ear isolates with that of commensal isolates was 1.42 (95% CI (1.13, 1.92); that is, OM isolates were 1.42 times more likely to be *hmwA*-positive than were commensal isolates. When stratified by region, *hmwA* was significantly more prevalent among OM isolates than commensal isolates within the U. S. ($P = 0.042$) and Finnish ($P = 0.048$) collections.

3.2. *hmwA* locus-specific screening

First, all 103 *hmwA*-leader positive isolates were interrogated with a combination of Southern blotting and locus-specific PCR to determine if they possessed two *hmw* loci. Ninety-eight (95.1%) of the *hmwA*-leader positive isolates contained two *hmwA* loci. There

were five isolates, however, in which the *hmwA* probe hybridized to a single genomic region when analyzed by Southern blotting, suggestive of a single *hmwA* locus. Alternatively, these five isolates may possess two *hmw* loci, but lack a *Bgl* II or *Eco* RI restriction site in the genomic region between the loci.

Next, using the locus-specific PCR assay results, we assessed whether the *hmw* loci were at the conserved chromosomal locations as described by Buscher et al., (Buscher et al., 2004); *i.e.*, adjacent to Rd ORFs HI1679 and HI1598. In these assays, the 5' PCR primer annealed to a 3'-region of the gene upstream of each *hmw* locus and the 3' PCR primer annealed within the *hmw* locus (Fig. 1); thus, locus number and chromosomal location was determined simultaneously. Eighty-four isolates (81.6%) contained *hmwA* loci located in the conserved locations described by Buscher et al., (Buscher et al., 2004).

Nineteen *hmwA*-positive isolates were locus-specific PCR-negative for one, or both, *hmw* loci. Thus, *hmw* locus number was assessed in these isolates by Southern blotting. Fourteen of these isolates possessed two genomic regions that hybridized to the *hmwA* probe, but when subjected to locus-specific PCR assays, only a single locus was detected. Six of these isolates were PCR-positive for *hmw*₁₆₇₉ only, one was positive for *hmw*₁₅₉₈ only, and seven were PCR-negative for both loci (Table S1). Failure to detect one, or both, loci by locus-specific PCR, would suggest that they are located in novel chromosomal locations, or that sequence variability precluded PCR amplification.

Lastly, five isolates contained a single *hmw* locus as detected by Southern blotting, four were PCR-positive for *hmwA*₁₆₇₉ and one isolate was PCR-negative for both loci (Table S1). Detection of a single *hmwA* band via Southern blotting, coupled with PCR amplification of a single locus, is consistent with an isolate possessing only a single *hmw* locus; these results, however, can also be explained by lack of restriction sites, loci at novel chromosomal locations, and/or nucleotide diversity in the primer annealing regions.

3.3. *hmwA* sequencing

The *hmwA*₁₆₇₉ and *hmwA*₁₅₉₈ binding regions were sequenced from 33 isolates selected for their diverse phylogenetic relationships estimated by MLST analysis (Fig. 2) (Lacross et al., 2013). First, by selecting isolates distributed across the MLST phylogeny, it was our intention to capture a wide range of *hmwA* diversity. Of the 33 isolates, 12 were from Finland, 11 from Israel, 10 from the US. Secondly, given the current lack of *hmwA* sequence data for commensal isolates, we wanted to include approximately equal numbers of OM and commensal isolates; in total, 17/33 isolates used for binding sequence analysis were OM isolates (Table S2).

The *hmwA* binding regions ranged in size from 2652 to 3573 nucleotides (Table S2), which is consistent with previous studies (Buscher et al., 2004; Dawid et al., 2001; Giufre et al., 2008; St Geme et al., 1998; van Schilfgaarde et al., 2000). Sequence data were obtained for both *hmwA*₁₆₇₉ and *hmwA*₁₅₉₈ from 26/33 of the isolates chosen for sequence analysis and for the remaining seven isolates *hmwA* sequence was obtained for one of the two loci (Table S2). Five of the seven *hmwA* sequences were excluded due to poor quality sequence and two additional *hmwA* sequences were excluded because they encoded in-frame stop codons.

Pairwise evolutionary distances estimated from the *hmwA* nucleotide sequence data ranged from 0.00 to 1.29 base substitutions per site with an overall average genetic distance of 0.7531 (SE = 0.04327). Pairwise *hmwA* amino acid p-distances, that is, the proportion of amino acids that vary between two sequences, ranged from 0.0000 to 0.4898 and the overall average pairwise amino acid distance was 0.3838 (SE = 0.0093), indicating that, on average, the *hmwA* binding region sequences shared approximately 62% amino acid identity.

Four pairs of *hmwA* binding region sequences were 100% identical at the nucleotide level (F2865 *hmwA*₁₆₇₉ and F894 *hmwA*₁₅₉₈ ; I280 *hmwA*₁₆₇₉ and I280 *hmwA*₁₅₉₈ ; K15RE *hmwA*₁₅₉₈ and H0421 *hmwA*₁₆₇₉ ; I283 *hmwA*₁₅₉₈ and G423 *hmwA*₁₅₉₈), but only strain I280 contained identical *hmwA* sequences at both loci. When considering the translated *hmwA* nucleotide sequences, four isolates (I280, F651, F1015, and F11242) possessed binding regions at both loci that varied by no more than three amino acids within an isolate; thus, four isolates encoded identical or nearly identical HMW adhesin binding at both of their *hmw* loci.

3.4. *hmwA* evolutionary analyses

The PHI test (Bruen et al., 2006) detected evidence of recombination ($P < 0.001$) within the *hmwA* sequences. Since recombination can preclude representing phylogenetic relationships among sequences with a single phylogeny, we employed a network approach to visualize *hmwA* relationships (Fig. S1). The Neighbor-Net LSfit index was 0.993, meaning that the phenetic distances between sequences on the network captured 99.3% of the estimated maximum composite likelihood genetic distances. The network topology (Fig. S1) was identical to the bootstrap consensus tree estimated by the maximum likelihood method (Fig. 3) using maximum composite likelihood distances; thus, despite the evidence for recombination, the dataset is still relatively tree-like.

The *hmwA* phylogeny revealed four *hmwA* sequence clusters, but the majority of *hmwA* sequences (83%) fell into either Clusters 1 or 2 (Fig. 3). Interestingly, of the 31 isolates for which both *hmwA* loci were sequenced, 29 (94%) possessed at least one locus that belonged to Cluster 1 (Fig. 3 and S2). In two isolates (F651 and I280) both genes, *hmwA*₁₆₇₉ and *hmwA*₁₅₉₈, were placed in Cluster 1; in one isolate (F1015) both genes were placed in Cluster 4; and in one isolate (F1124-2) both genes were placed in Cluster 2.

Evolutionary distances and amino acid p-distances were estimated for Cluster 1 and Cluster 2 sequences separately whereas Clusters 3 and 4 sequences were combined into a single cluster for analysis (referred to as Cluster 3–4). The average evolutionary distances ranged from 0.502 to 0.636 nucleotide substitutions per site for Cluster 3–4 and Cluster 1, respectively. Average amino acid p-distances ranged from 0.3341 for Cluster 3–4 to 0.3467 for Cluster 1 suggesting that, on average, any two Cluster 1 sequences differ at approximately 35% of their amino acid sites (data not shown).

hmwA sequences did not cluster by chromosomal locus, geographic region, or by disease status (OM or commensal). There are, however, OM and commensal isolates of the same MLST sequence type (ST) that also have closely related *hmwA* sequences at both loci (Fig. 3 and Table S2), for example, H0421 and K15RE2.6 are both ST-11, F1646 and F1942 are

both ST-3, F894 and F2865 are both ST-155, and F1296-5 and K19RE2.4 are both ST-57. With the exception of the ST-57 isolates, each ST pair with closely related *hmwA* loci consists of one OM and one commensal isolate collected from the same geographic region. The two ST-57s with similar *hmwA* sequences at both loci are both OM isolates, but, one was isolated in Finland and the other in the U. S.

4. Discussion

In this study, we determined the prevalence of *hmwA* in a collection of 170 well characterized NTHi isolates (Lacross et al., 2013) collected between 1994 and 2002 from the throats of healthy children and the middle ears of children with acute otitis media from three disparate geographic regions. Overall, *hmwA* prevalence was 61% and otitis media isolates were significantly more likely to possess *hmwA* (71%) than were commensal isolates (48%). To gain insight into *hmwA* sequence diversity, we sequenced the binding regions of both *hmwA* loci (*hmwA*₁₆₇₉ and *hmwA*₁₅₉₈) from 33 isolates, documenting a wide range of nucleotide diversity within and between isolates. Phylogenetic analyses of the *hmwA* binding regions revealed four distinct sequence clusters, but, sequences did not cluster by geographic region, *hmwA* locus, or disease status.

The *hmwA* prevalence (61%) in this isolate collection was consistent with previous studies, which report prevalences ranging from 38 – 80% (Barenkamp, 1992; Ecevit et al., 2004; Erwin et al., 2005; Erwin et al., 2008; St Geme et al., 1998; van Schilfgaarde et al., 2000). The wide ranging reported prevalence likely reflects a combination of sample composition and detection methods (Barenkamp and Leininger, 1992; Ecevit et al., 2004; Erwin et al., 2005; Erwin et al., 2008; St Geme et al., 1998; van Schilfgaarde et al., 2000). When stratified by geographic region, isolates collected from Finland were significantly more likely to be *hmwA*-positive than were those from Israel or the U.S. (Table 2). The reasons for increased *hmwA* prevalence among Finnish isolates is not readily apparent, and further analyses are necessary to determine if this result reflects true differences in the geographic distribution of *hmwA*. Otitis media isolates were significantly more likely to possess *hmwA* than were commensal isolates across all regions and within each geographic region. Similar associations between the presence of *hmwA*, or HMW-adhesins, and disease status have been previously reported (Ecevit et al., 2004; van Schilfgaarde et al., 2000), suggesting an important role in for HMW adhesins in NTHi pathogenesis. Previous studies, however, have demonstrated that OM isolates produce less HMW adhesin than their isogenic counterparts collected at the same time from the throat (Dawid et al., 1999). Thus, it is unclear if HMW adhesins play a direct role in OM pathogenesis or if their association with disease is due to linkage with other virulence determinants.

Characterization of the isolates in this collection by multi-locus sequence-typing (MLST) identified 109 different sequence types (STs), with 28 STs represented by two or more isolates, and detected a clear signal of population structure (Lacross et al., 2013). MLST-based phylogenetic analysis revealed that these NTHi isolates do not cluster by disease status or by geographic region (Lacross et al., 2013), and mapping of *hmwA* status onto the phylogeny showed that *hmwA*-positive isolates were distributed throughout the phylogeny (Fig. 2). This wide distribution of *hmwA* across phylogenetic clusters is consistent with

previous studies based on multi-locus enzyme electrophoresis (MLEE) (St Geme et al., 1998) and MLST (Erwin et al., 2008). However, ST was a strong predictor of *hmwA* status in our isolate collection—of the 28 STs represented more than once, isolates from 26 STs were concordant for the presence, or absence, of *hmwA*. A more recent analysis of 97 NTHi isolates, based upon whole genome sequences, identified six populations that correspond to distinct phylogenetic clades (De Chiara et al., 2014). Similar to the MLST based study by Lacross, et al. (Lacross et al., 2013), there was little evidence of clustering by geographic origin of the isolate or by disease status (e.g., otitis media, COPD, etc.) (De Chiara et al., 2014). Furthermore, there were strong associations between the presence, or absence, of *hmw* loci and population structure in both collections. For example, mapping *hmwA*-status to populations as defined in Lacross et al. (Lacross et al., 2013), reveals that among isolates with > 80% of their ancestry attributable to population 1, 13/14 are *hmwA*-negative, whereas isolates with > 80% of their ancestry attributable to populations 6 (n = 13), 7 (n = 18) and 8 (n = 5) are universally *hmwA*-positive (see Table S1 of Lacross et al. (Lacross et al., 2013) for isolate ancestry). Similarly, *hmw* loci were significantly associated with four of the six NTHi clades identified by De Chiara, et al. (De Chiara et al., 2014). Thus, our findings lend further support to the idea that *hmwA* loci are largely maintained within clonal NTHi lineages.

Ninety-five percent (98/103) of the *hmwA*-positive isolates possessed two *hmwA* loci, and in 87% of those isolates, the loci were present at the conserved chromosomal locations identified by Buscher et al. (Buscher et al., 2004). The majority of the remaining *hmwA*-positive isolates contained two *hmwA* loci as detected by Southern blotting, but locus-specific PCR amplification was unsuccessful for at least one locus, raising the possibility that the *hmwA* loci can reside in alternative chromosomal locations. Interestingly, 12/14 of the isolates with two loci but negative locus-specific PCR results were OM isolates. The observation that *hmw* loci are nearly always present in two copies is interesting given that they are not genetically linked (Buscher et al., 2004). Furthermore, the loci usually differ in their amino acid sequences and confer differing adherence specificities (Buscher et al., 2004; St Geme, 1994; St Geme et al., 1993; van Schilfgaarde et al., 2000). Taken together, this suggests that having two different HMW adhesins confers an important fitness advantage relative to having a single adhesin or two copies of the same adhesin. This advantage could be realized through a combination of niche expansion, immune evasion, and/or increased adherence.

Targeted sequencing of the *hwmA* binding regions from 33 isolates (Table S4 and Fig. 2) revealed a wide range of diversity in size and amino acid composition. The observed size and amino acid diversity of the binding region is consistent with other studies (Barenkamp and Bodor, 1990; Barenkamp and Leininger, 1992; Buscher et al., 2004; Cholon et al., 2008; Dawid et al., 2001; Ecevit et al., 2005; Giufre et al., 2008; Giufre et al., 2006; van Schilfgaarde et al., 2000). Five isolates, however, possessed identical or nearly identical *hmwA* sequences at both *hmw* loci suggestive of recent gene conversion events such as those described among outer membrane proteins of the gastric pathogen *Helicobacter pylori* (Solnick et al., 2004; Talarico et al., 2012). A NTHi isolate that encodes two nearly identical HMW adhesins may be restricted to a narrowly defined environmental niche but display

increased adherence within that niche. Furthermore, an isolate with limited HMW adhesin diversity may incur an increased fitness cost, relative to an isolate with two immunologically distinct adhesins, in the presence of antibody-mediated immunity; however, fitness costs could potentially be mitigated by simple sequence repeat mediated *hmwA* phase variation. It should be noted, however, that the four nearly identical *hmwA* sequences described here differ at a small number of amino acid sites and these minor differences could potentially affect adherence properties and/or immunogenicity of the encoded adhesin.

Phylogenetic analyses of the *hmwA* binding regions revealed four distinct sequence clusters, with the majority of sequences (84%) falling into either Cluster 1 or Cluster 2. There was no evidence for OM-specific *hmwA* sequences—each sequence cluster contained *hmwA* sequences from both OM and commensal isolates. Consistent with the view that OM isolates represent a limited subset of the much larger population of commensal isolates, commensal isolates from Cluster 1 and Cluster 2 demonstrated a wider range of genetic diversity than OM isolates.

The HMW adhesins of NTHi strain 12 are known to differ in their *in vitro* adherence properties (Buscher et al., 2004; St Geme, 1994; St Geme et al., 1993; van Schilfgaarde et al., 2000). Four distinct *in vitro* adherence patterns were identified among a diverse collection of NTHi isolates (van Schilfgaarde et al., 2000), which is interesting given that we identified four *hmwA* sequence clusters. Later studies demonstrated that *hmwA* sequences cluster by their *in vitro* adherence properties (Buscher et al., 2004); for example, NTHi strain 12 *hmwA*₁₆₇₉ and strain 5 *hmwA*₁₅₉₈ both fall within Cluster 1 (Fig 3) and have Strain 12 HMW2-like adherence properties (Buscher et al., 2004). Assuming phylogenetic relatedness of DNA sequences reflects functional characteristics of the encoded adhesin, Cluster 1 sequences would be expected to exhibit NTHi strain 12 HMW2-like binding, which is included in Cluster 1. Cluster 1 sequences are particularly interesting because nearly all of the isolates sequenced (94%) possessed a *hmwA* that fell within phylogenetic Cluster 1. This observation raises the interesting possibility that Cluster 1 sequences confer an ancestral binding functionality that is of primary importance and that following duplication, the second *hmwA* has taken on additional adherence properties.

This study has greatly increased the number of publicly available *hmwA* sequences, especially those from commensal NTHi isolates. Our sample size, however, was relatively small and we have likely only scratched the surface of *hwmA* diversity. Thus, as more isolates are sequenced, additional sequence clusters may be identified. It is also important to recognize the limitations of assessing the role of HMW adhesins during OM pathogenesis based solely on prevalence data since the presence of *hmwA* illuminates the genetic potential of a strain but provides limited insight regarding the expression of a functional HMW adhesin. The relationship between *hwmA* presence and HMW expression is complicated by the fact that *hmwA* is phase variable. Several studies have demonstrated that phase variable HMW adhesin expression has important implications for NTHi pathogenesis (Cholon et al., 2008; Davis et al., 2014; Dawid et al., 1999). Increased *hmwA* prevalence among OM isolates compared to commensal isolates suggests that *hmwA*-containing isolates are more likely than their *hmwA*-negative counterparts to survive in the middle ear space, and thus supports a role for HMW adhesins in NTHi virulence.

Future studies aimed at defining the adherence profiles of Cluster 3 and 4 sequences will help determine if these proteins display novel adherence characteristics. Finally, this work has important public health implications since the HMW adhesins have been proposed as potential components of a NTHi targeted vaccine (Winter and Barenkamp, 2014). Phylogenetic analysis can be used to inform the selection of HMW adhesins for inclusion in a vaccine, e.g., selecting representative alleles from each phylogenetic cluster. At the same time, the extensive amino acid diversity observed in this collection, and the lack of disease-specific *hmwA* sequence clusters, underscores the challenges of targeting HMW adhesins as vaccine components.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported with funding from the Interdisciplinary Program in Infectious Diseases (TA32A1049816), the Molecular Mechanisms of Microbial Pathogenesis Training Program (AI007528), and the University of Michigan, School of Public Health, Department of Epidemiology. This research was additionally supported with National Institute of Health (R01 DC005840 and R01 HL083893) funding awarded to JRG.

Literature cited

- Barenkamp SJ. Outer membrane proteins and lipopolysaccharides of nontypeable *Haemophilus influenzae*. *J Infect Dis*. 1992; 165(Suppl 1):S181–S184. [PubMed: 1588159]
- Barenkamp SJ, Bodor FF. Development of serum bactericidal activity following nontypable *Haemophilus influenzae* acute otitis media. *Pediatr Infect Dis J*. 1990; 9:333–339. [PubMed: 2352818]
- Barenkamp SJ, Leininger E. Cloning, expression, and DNA sequence analysis of genes encoding nontypeable *Haemophilus influenzae* high-molecular-weight surfaceexposed proteins related to filamentous hemagglutinin of *Bordetella pertussis*. *Infect Immun*. 1992; 60:1302–1313. [PubMed: 1548058]
- Barenkamp SJ, St Geme JW 3rd. Genes encoding high-molecular-weight adhesion proteins of nontypeable *Haemophilus influenzae* are part of gene clusters. *Infect Immun*. 1994; 62:3320–3328. [PubMed: 8039903]
- Bruen TC, Philippe H, Bryant D. A simple and robust statistical test for detecting the presence of recombination. *Genetics*. 2006; 172:2665–2681. [PubMed: 16489234]
- Bryant D, Moulton V. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol*. 2004; 21:255–265. [PubMed: 14660700]
- Buscher AZ, Burmeister K, Barenkamp SJ, St Geme JW 3rd. Evolutionary and functional relationships among the nontypeable *Haemophilus influenzae* HMW family of adhesins. *J Bacteriol*. 2004; 186:4209–4217. [PubMed: 15205423]
- Cholon DM, Cutter D, Richardson SK, Sethi S, Murphy TF, Look DC, St Geme JW 3rd. Serial isolates of persistent *Haemophilus influenzae* in patients with chronic obstructive pulmonary disease express diminishing quantities of the HMW1 and HMW2 adhesins. *Infect Immun*. 2008; 76:4463–4468. [PubMed: 18678658]
- Davis GS, Marino S, Marrs CF, Gilsdorf JR, Dawid S, Kirschner DE. Phase variation and host immunity against high molecular weight (HMW) adhesins shape population dynamics of nontypeable *Haemophilus influenzae* within human hosts. *J Theor Biol*. 2014; 355:208–218. [PubMed: 24747580]

- Dawid S, Barenkamp SJ, St Geme JW 3rd. Variation in expression of the *Haemophilus influenzae* HMW adhesins: a prokaryotic system reminiscent of eukaryotes. *Proc Natl Acad Sci U S A*. 1999; 96:1077–1082. [PubMed: 9927696]
- Dawid S, Grass S, St Geme JW 3rd. Mapping of binding domains of nontypeable *Haemophilus influenzae* HMW1 and HMW2 adhesins. *Infect Immun*. 2001; 69:307–314. [PubMed: 11119519]
- De Chiara M, Hood D, Muzzi A, Pickard DJ, Perkins T, Pizza M, Dougan G, Rappuoli R, Moxon ER, Soriani M, Donati C. Genome sequencing of disease and carriage isolates of nontypeable *Haemophilus influenzae* identifies discrete population structure. *Proc Natl Acad Sci U S A*. 2014; 111:5439–5444. [PubMed: 24706866]
- Ecevit IZ, McCrea KW, Marrs CF, Gilsdorf JR. Identification of new hmwA alleles from nontypeable *Haemophilus influenzae*. *Infect Immun*. 2005; 73:1221–1225. [PubMed: 15664969]
- Ecevit IZ, McCrea KW, Pettigrew MM, Sen A, Marrs CF, Gilsdorf JR. Prevalence of the *hifBC*, *hmw1A*, *hmw2A*, *hmwC*, and *hia* Genes in *Haemophilus influenzae* Isolates. *J Clin Microbiol*. 2004; 42:3065–3072. [PubMed: 15243061]
- Erwin AL, Nelson KL, Mhlanga-Mutangadura T, Bonthuis PJ, Geelhood JL, Morlin G, Unrath WC, Campos J, Crook DW, Farley MM, Henderson FW, Jacobs RF, Muhlemann K, Satola SW, van Alphen L, Golomb M, Smith AL. Characterization of genetic and phenotypic diversity of invasive nontypeable *Haemophilus influenzae*. *Infect Immun*. 2005; 73:5853–5863. [PubMed: 16113304]
- Erwin AL, Sandstedt SA, Bonthuis PJ, Geelhood JL, Nelson KL, Unrath WC, Diggle MA, Theodore MJ, Pleatman CR, Mothershed EA, Sacchi CT, Mayer LW, Gilsdorf JR, Smith AL. Analysis of genetic relatedness of *Haemophilus influenzae* isolates by multilocus sequence typing. *J Bacteriol*. 2008; 190:1473–1483. [PubMed: 18065541]
- Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*. 1985; 39:783–791.
- Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*. 1995; 269:496–512. [PubMed: 7542800]
- Giufre M, Carattoli A, Cardines R, Mastrantonio P, Cerquetti M. Variation in expression of HMW1 and HMW2 adhesins in invasive nontypeable *Haemophilus influenzae* isolates. *BMC Microbiol*. 2008; 8:83. [PubMed: 18510729]
- Giufre M, Muscillo M, Spigaglia P, Cardines R, Mastrantonio P, Cerquetti M. Conservation and diversity of HMW1 and HMW2 adhesin binding domains among invasive nontypeable *Haemophilus influenzae* isolates. *Infect Immun*. 2006; 74:1161–1170. [PubMed: 16428765]
- Grass S, Buscher AZ, Swords WE, Apicella MA, Barenkamp SJ, Ozchlewski N, St Geme JW 3rd. The *Haemophilus influenzae* HMW1 adhesin is glycosylated in a process that requires HMW1C and phosphoglucomutase, an enzyme involved in lipooligosaccharide biosynthesis. *Mol Microbiol*. 2003; 48:737–751. [PubMed: 12694618]
- Grass S, St Geme JW 3rd. Maturation and secretion of the non-typable *Haemophilus influenzae* HMW1 adhesin: roles of the N-terminal and C-terminal domains. *Mol Microbiol*. 2000; 36:55–67. [PubMed: 10760163]
- Hogg JS, Hu FZ, Janto B, Boissy R, Hayes J, Keefe R, Post JC, Ehrlich GD. Characterization and modeling of the *Haemophilus influenzae* core and supragenomes based on the complete genomic sequences of Rd and 12 clinical nontypeable strains. *Genome Biol*. 2007; 8:R103. [PubMed: 17550610]
- Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*. 2006; 23:254–267. [PubMed: 16221896]
- Kong F, Brown M, Sabananthan A, Zeng X, Gilbert GL. Multiplex PCR-based reverse line blot hybridization assay to identify 23 *Streptococcus pneumoniae* polysaccharide vaccine serotypes. *J Clin Microbiol*. 2006; 44:1887–1891. [PubMed: 16672432]
- Lacross NC, Marrs CF, Gilsdorf JR. Population Structure in Nontypeable *Haemophilus influenzae*. *Infect Genet Evol*. 2013
- Loytynoja A, Goldman N. webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinformatics*. 2010; 11:579. [PubMed: 21110866]

- Nei, M.; Kumar, S. *Molecular Evolution and Phylogenetics*. New York: Oxford University Press; 2000.
- R Core Team. R: A Language and Environment for Statistical Computing, 3.0.1 ed. Vienna, Austria: R Foundation for Statistical Computing; 2014.
- Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques*. 2003; 34:374–378. [PubMed: 12613259]
- Sandstedt SA, Zhang L, Patel M, McCrea KW, Qin Z, Marrs CF, Gilsdorf JR. Comparison of laboratory-based and phylogenetic methods to distinguish between *Haemophilus influenzae* and *H. haemolyticus*. *J Microbiol Methods*. 2008; 75:369–371. [PubMed: 18652852]
- Solnick JV, Hansen LM, Salama NR, Boonjakuakul JK, Syvanen M. Modification of *Helicobacter pylori* outer membrane protein expression during experimental infection of rhesus macaques. *Proc Natl Acad Sci U S A*. 2004; 101:2106–2111. [PubMed: 14762173]
- St Geme JW 3rd. The HMW1 adhesin of nontypeable *Haemophilus influenzae* recognizes sialylated glycoprotein receptors on cultured human epithelial cells. *Infect Immun*. 1994; 62:3881–3889. [PubMed: 8063405]
- St Geme JW 3rd, Cutter D. Evidence that surface fibrils expressed by *Haemophilus influenzae* type b promote attachment to human epithelial cells. *Mol Microbiol*. 1995; 15:77–85. [PubMed: 7752898]
- St Geme JW 3rd, Cutter D. Influence of pili, fibrils, and capsule on in vitro adherence by *Haemophilus influenzae* type b. *Mol Microbiol*. 1996; 21:21–31. [PubMed: 8843431]
- St Geme JW 3rd, de la Morena ML, Falkow S. A *Haemophilus influenzae* IgA protease-like protein promotes intimate interaction with human epithelial cells. *Mol Microbiol*. 1994; 14:217–233. [PubMed: 7830568]
- St Geme JW 3rd, Falkow S, Barenkamp SJ. High-molecular-weight proteins of nontypeable *Haemophilus influenzae* mediate attachment to human epithelial cells. *Proc Natl Acad Sci U S A*. 1993; 90:2875–2879. [PubMed: 8464902]
- St Geme JW 3rd, Grass S. Secretion of the *Haemophilus influenzae* HMW1 and HMW2 adhesins involves a periplasmic intermediate and requires the HMWB and HMWC proteins. *Mol Microbiol*. 1998; 27:617–630. [PubMed: 9489673]
- St Geme JW 3rd, Kumar VV, Cutter D, Barenkamp SJ. Prevalence and distribution of the *hmw* and *hia* genes and the HMW and Hia adhesins among genetically diverse strains of nontypeable *Haemophilus influenzae*. *Infect Immun*. 1998; 66:364–368. [PubMed: 9423882]
- Swords WE, Buscher BA, Ver Steeg Ii K, Preston A, Nichols WA, Weiser JN, Gibson BW, Apicella MA. Non-typeable *Haemophilus influenzae* adhere to and invade human bronchial epithelial cells via an interaction of lipooligosaccharide with the PAF receptor. *Mol Microbiol*. 2000; 37:13–27. [PubMed: 10931302]
- Talarico S, Whitefield SE, Fero J, Haas R, Salama NR. Regulation of *Helicobacter pylori* adherence by gene conversion. *Mol Microbiol*. 2012; 84:1050–1061. [PubMed: 22519812]
- Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci U S A*. 2004; 101:11030–11035. [PubMed: 15258291]
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011; 28:2731–2739. [PubMed: 21546353]
- van Schilfgaarde M, van Ulsen P, Eijk P, Brand M, Stam M, Kouame J, van Alphen L, Dankert J. Characterization of adherence of nontypeable *Haemophilus influenzae* to human epithelial cells. *Infect Immun*. 2000; 68:4658–4665. [PubMed: 10899870]
- Winter LE, Barenkamp SJ. Antibodies to the HMW1/HMW2 and Hia Adhesins of Nontypeable *Haemophilus influenzae* Mediate Broad-Based Opsonophagocytic Killing of Homologous and Heterologous Strains. *Clin Vaccine Immunol*. 2014; 21:613–621. [PubMed: 24574538]
- Yang Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol*. 1994; 39:306–314. [PubMed: 7932792]

Zhang L, Srinivasan U, Marrs CF, Ghosh D, Gilsdorf JR, Foxman B. Library on a slide for bacterial comparative genomics. *BMC Microbiol.* 2004; 4:12. [PubMed: 15035675]

Highlights

- *hmwA* was significantly more prevalent among otitis media NTHi than commensal NTHi
- Ninety-five percent (98/103) of the *hmwA*-positive isolates possessed two *hmw* loci
- Phylogenetic analyses of *hmwA* sequences revealed four distinct sequence clusters
- *hmwA* sequences did not cluster by chromosomal location, geography, or disease

Figure 1.

Schematic representation of primer annealing regions relative to a generalized *hmwIA* locus.

(a) Each NTHi encodes two *hmw* loci, consisting of *hmwA*, *hmwB*, and *hmwC*, located at conserved, genetically unlinked, chromosomal regions (Buscher et al., 2004). The functional HMW adhesin is encoded by *hmwA*, which includes conserved signal sequences that are cleaved during HMW maturation (Barenkamp and Leininger, 1992). (b) Because the signal sequences are conserved, we targeted this region to estimate *hmwA* prevalence using both PCR and probe based methods. (c) To determine the number and chromosomal location of each *hmwA* locus, PCR primers that annealed to genes upstream of each *hmwA* locus were utilized along with a reverse primer that recognized a conserved region of *hmwA*. Next, nearly full-length *hmwA* mature-protein regions were amplified and this served as template for PCR amplification of the binding domain region that was sequenced.

Figure 2.

Unrooted phylogeny estimated from MLST data for the 170 isolates used in this study (phylogeny taken from Fig. 3 (Lacross et al. 2013)). Columns one and two are color coded according to the geographic region from which the strain was collected (I) and the disease status of the individual (II) from which the strain was isolated. All 103 *hmrA* leader positive isolates (III) were interrogated by locus-specific PCR assays targeting *hmrA*₁₆₇₉ (IV) and *hmrA*₁₅₉₈ (V). A subset of isolates (n = 22) were screened for *hmrA* by Southern blotting, 19 of these isolates were *hmrA*-positive by Southern blotting, but PCR-negative for one, or

both, *hmw* loci (VI). Isolates chosen for *hmwA* binding region DNA sequence analysis are indicated with red arrows.

Figure 3. *hmwA* core binding region evolutionary relationships were estimated using the maximum likelihood approach. *hmwA* core binding region sequences formed four distinct sequence clusters and most isolates (94%) possessed a *hmwA* that fell within Cluster 1. The sequences did not cluster by *hmwA* locus (I), by geographic region (II), or by disease status (III). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein, 1985); branches recovered in less than 50% of the bootstrap replicates were collapsed. Differences in the evolutionary

rates among sites were modeled with a five-category discrete gamma distribution (shape parameter = 0.6044). All positions containing gaps and missing data were eliminated from the analysis resulting in a total of 1809 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura et al., 2011).

Table 1

Polymerase chain reaction (PCR) primers used in this study.

Primer Name	PCR Primer Sequence (5' to 3')	Reference
<i>hmwA_leader_F</i>	GCCAATTTCCGCTTCACCCTCTTT	this study
<i>hmwA_leader_R</i>	AACAACCTCCGCCGTATTTAACCGC	this study
A1FL5.2	YGATAGSGTAGATCTCCCCGCCTTTGC	this study
<i>hmwA1F_ORF5-1</i>	TGGAACCTCTTTTGCTGTGGCTGATGC	(Giufre et al., 2006)
A2FL5.2	TTGTGGCAATTCAATACCTATTTGTGG	this study
<i>hmwA2F_ORF5-1</i>	CCTCTTAATTGGGCATTAGTTGG	(Giufre et al., 2006)
<i>hmwA_core_1R</i>	CCGGTGATATTCACGCTGCTTGAGG	this study
<i>hmwAR_HMW3R</i>	GATGAAGAAGCCAGGCCAAGCAATAC	(Giufre et al., 2006).
<i>hmwAseqR_NotI^a</i>	ATGATCAGCGCCGCGSGTGATRTTYACDYTRCTTGAGG	this study
<i>hmwAseqF_SpeI^a</i>	CAGTCAGACTAGTAAGGYAAAAAYGGYATTCAATTAGC	this study

^a restriction enzyme recognition site

Table 2*hmwA* prevalence.

All	<i>hmwA</i> ⁺ (%)	<i>hmwA</i> ⁻ (%)	PR (95% CI) ^a	P-value ^b
OM (n = 95)	67 (71)	28 (29)	1.47 (1.12, 1.92)	0.004
Commensal (n = 75)	36 (48)	39 (52)		
Total (n = 170)	103 (61)	67 (39)		
Finland				
OM (n = 29)	26 (90)	3 (10)	1.37 (1.01, 1.86)	0.048
Commensal (n = 26)	17 (65)	9 (35)		
Total (n = 55)	43 (78)	12 (22)		
Israel				
OM (n = 28)	15 (54)	13 (46)	1.47 (0.77, 2.83)	0.264
Commensal (n = 22)	8 (36)	14 (64)		
Total (n = 50)	23 (46)	27 (54)		
U. S.				
OM (n = 38)	26 (68)	12 (32)	1.68 (1.01, 2.78)	0.042
Commensal (n = 27)	11 (41)	16 (59)		
Total (n = 65)	37 (57)	28 (43)		

^aPrevalence ratio (OM/commensal) and 95% confidence intervals^bFisher's exact test two-tailed P-value