bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 # Characterisation of the phase-variable autotransporter Lav reveals a role in host cell
2 # adherence and biofilm formation in Non-Typeable *Haemophilus influenzae*

3 Zachary N. Phillips[1], Preeti Garai[2], Greg Tram[1], Asma-Ul Husna[1], Megan Staples[3], Keith
4 Grimwood[4], Amy V. Jennison[3], Michael P. Jennings[1], Kenneth L. Brockman[2] & John M.
5 Atack[1]*

6 [1]Institute for Glycomics, Griffith University, Gold Coast, QLD 4215, Australia.

7 [2]Department of Microbiology & Immunology, Medical College of Wisconsin, Milwaukee,
8 Wi, USA

9 [3]Queensland Department of Health, Public Health Microbiology, Forensic and Scientific
10 Services, Brisbane, Queensland, Australia

11 [4]School of Medicine and Dentistry and Menzies Health Institute Queensland, Griffith
12 University, and Departments of Infectious Diseases and Paediatrics, Gold Coast Health,
13 Queensland, Australia

14

15 *Corresponding author: John M Atack. Institute for Glycomics, Griffith University, Gold
16 Coast, QLD 4215, Australia. Email: j.atack@griffith.edu.au; Tel: +61 07 567 80580; Fax:
17 +61 07 555 28098

18

19 **Key words:** *Lav, autotransporter, phase variation, adherence, biofilm, Haemophilus*

20

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

21 **Abstract**

22 Lav is an autotransporter protein found in pathogenic *Haemophilus* and *Neisseria* species.
23 Lav in non-typeable *Haemophilus influenzae* (NTHi) is phase-variable: the gene reversibly
24 switches ON-OFF via changes in length of a locus-located $GCAA_{(n)}$ simple DNA sequence
25 repeat tract. The expression status of *lav* was examined in carriage and invasive collections of
26 NTHi, where it was predominantly not expressed (OFF). Phenotypic study showed *lav*
27 expression (ON) results in increased adherence to host cells, and denser biofilm formation. A
28 survey of *Haemophilus* spp. genome sequences showed *lav* is present in ~60% of NTHi
29 strains, but *lav* is not present in most typeable *H. influenzae*. Sequence analysis revealed a
30 total of five distinct variants of the Lav passenger domain present in *Haemophilus* spp., with
31 these five variants showing a distinct lineage distribution. Determining the role of Lav in
32 NTHi will help understand the role of this protein during distinct pathologies.

33

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

34 **Introduction**

35 Non-typeable *Haemophilus influenzae* (NTHi) is a bacterial pathogen of global importance.
36 NTHi colonizes the human nasopharynx, but is an important pathogen in middle ear infection
37 (otitis media) in children (1), exacerbations in bacterial bronchitis, chronic obstructive
38 pulmonary disease and bronchiectasis (2, 3), and community-acquired pneumonia, in adults
39 (4). NTHi also causes invasive infections, and these are fatal in ~10% of children <1 year,
40 and in ~25% of adults aged ≥80-years (5-7). Frequency of disease caused by NTHi is
41 increasing annually, exacerbated by both the absence of an NTHi vaccine, and by emerging
42 antibiotic-resistance (8). Understanding pathobiology, and identifying the stably expressed
43 antigenic repertoire, of NTHi is crucial for the rational design of a protein subunit vaccine,
44 but is complicated by factors like variable gene expression and low sequence conservation.

45 Several host-adapted bacterial pathogens are able to randomly and reversibly switch gene
46 expression, a process known as phase-variation (9-11). Many bacterial genes phase-vary by
47 changes in length of locus-located simple DNA sequence repeat (SSR) tracts. When SSR
48 tracts are located in the open reading frame of a gene, this variation in length results in ON-
49 OFF switching of expression. Phase-variable genes typically encode surface proteins such as
50 iron acquisition factors (11), lipooligosaccharide biosynthetic enzymes (12), and adhesins
51 (13). Phase-variation of bacterial surface features generates sub-populations of phenotypic
52 variants, some of which may be better adapted to a particular niche, or equipped to avoid an
53 immune response. Many bacterial surface proteins are classified as autotransporters, and
54 these contain a C-terminal β-barrel translocator domain in the outer membrane, and an
55 extracellular passenger domain (14). Many virulence-associated autotransporters are phase-
56 variably expressed, including UpaE in uropathogenic *Escherichia coli* (15), Hap (16) and Hia
57 (17) in *Haemophilus influenzae*, and NalP (18, 19), AutA (20) and AutB (21) in *Neisseria*
58 spp. A homologue of AutB, named Lav, has been described in multiple *Haemophilus* spp.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

59 (21, 22). The *lav* gene has also been reported to be phase-variable, as a $GCAA_{(n)}$ SSR tract is
60 present in the *lav* open reading frame (22). Investigation into AutB in *N. meningitidis* found
61 the protein played a role in biofilm formation, and was phase-varied OFF in available
62 genomes (21). Study of another Lav homologue, Las, in *H. influenzae* biogroup *aegyptius*,
63 has suggested a role in inflammatory cytokine production (23), and increased expression
64 associated with disease progression (24). The function of Lav in NTHi has not been studied
65 in detail, although over multiple rounds of infection the *lav* gene was shown to phase-vary
66 OFF (25), implying selection against Lav during chronic/recurrent infections. Therefore, we
67 sought to undertake a phenotypic characterisation of the role of Lav in NTHi, and to
68 determine the prevalence and diversity of this protein in *Haemophilus* spp.

69

70 **Materials & Methods**

71 ***Bacterial Isolate Collections.*** Nasal (carriage) control samples were taken from the ORChID
72 collection, a prospective birth cohort study of infants in South East Queensland. As part of
73 this collection, respiratory disease symptoms were recorded daily, and weekly nasal swabs
74 were collected, from 158 infants during their first two years of life (2010-2012) (26). All
75 samples used as carriage controls were randomly selected from infants demonstrating no
76 overt symptoms of respiratory illnesses either 2 weeks before or after sampling (27). Invasive
77 NTHi isolates used for this study were isolated from patients suffering from *H. influenzae*
78 infections in SE Queensland over a 15-year period (2001-2015) (28). Information on age,
79 sample site, and geographical location were collected, but not on comorbidities (28).

80 ***Bacterial growth and media.*** NTHi isolates were grown in brain heart infusion (BHI; Oxoid)
81 supplemented (sBHI) with hemin (1%) and β-NAD (2 μg/ml) at 37°C in an atmosphere
82 containing 5% (v/v) CO_2. *Escherichia coli* strains were grown in Luria-Bertani (LB) broth, or

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

83 on LB agar (LB broth +1.5% [w/v] bacteriological agar). LB was supplemented with
84 ampicillin (100µg/mL) as required.

85 ***SSR tract PCR and fragment analysis.*** Bacterial genomic DNA from invasive isolates was
86 prepared as described previously (29). Standard methods were used throughout for PCR
87 using GoTaq Flexi DNA polymerase according to manufacturer's instructions (Promega),
88 and fragment analysis was carried out as previously described (30). *lav* ON/OFF status was
89 determined from the number of GCAA repeats in the SSR tract present in the gene (based on
90 amplicon peak size), using Lav_F (FAM-GCCCCATTTATTTTTACTTGACAAAGG) and
91 Lav_R (GCTCATTTGTTAATTTAGAATTGTCATAAG) primers by sizing and quantifying
92 using the GeneScan system (Applied Biosystems International) at the Australian Genome
93 Research Facility (AGRF; Brisbane, Australia), and traces analysed using PeakScanner
94 software 2.0 (Applied Biosystems International). Enriched ON and OFF variants in strain 86-
95 028NP were generated by colony screening and enrichment for GCAA tract lengths in the *lav*
96 SSR tract.

97 ***Cloning Lav protein fragment for generation of antisera.*** The Lav passenger domain and
98 flanking region, comprising residues 250-540 of the full protein, was expressed by cloning
99 the encoding DNA into the pET15b vector, in-frame with the N-terminal His-tag, The coding
100 region was amplified from strain 86-028NP using primers Lav_bind-F
101 (AGTCAGCATATGCAAGATAACTCACACGTTATCG) and Lav_bind-R
102 (CTGACTGGATCCTTAGTGGCGGAAGCGTTGATATTG) with KOD HotStart
103 proofreading DNA polymerase (Novagen) and cloned into the NdeI and BamHI sites of
104 pET15b, to generate vector pET15b::Lav-bind. Expression was carried out using *E. coli*
105 BL21 (DE3) containing the pET15b::Lav-bind vector in LB broth induced with 1mM
106 Isopropyl β-d-1-thiogalactopyranoside (IPTG) at 37°C with shaking for 16 hours. Purification
107 with TALON Metal Affinity Resin (Takara) was carried out from the insoluble fraction by

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

108 using multiple rounds of sonication and washes in PBS containing 0.1% Tween (v/v).
109 Following purification, pure Lav-bind was dialysed at 4°C for 12 hours in PBS, twice.

110 ***Western Blotting.*** Protein lysates of whole NTHi were prepared by heating whole cell
111 suspensions at 99°C for 40 mins. These were electrophoresed on 4-12% Bis-Tris
112 polyacrylamide gels (Invitrogen) at 150V for 45mins in Bolt MOPS Running Buffer
113 (Invitrogen). Samples were transferred to nitrocellulose membrane at 15V for 1h. Membranes
114 were blocked with 5% (w/v) skim milk in Tris-buffered saline with 0.1% Tween-20 (TBS-T)
115 by shaking overnight at 4°C. Primary mouse antibodies against the Lav-bind protein (anti-
116 Lav antisera) were raised in BALB/c mice at the Institute for Glycomics Animal Facility.
117 50µg of purified Lav-bind protein in alum was used per mouse. Primary antibody was used at
118 1:1000 dilution in 5% (w/v) skim milk in TBS-T for 1h with shaking at room temperature.
119 Membranes were washed multiple times in TBS-T for 1h before adding secondary antibody
120 (goat anti-mouse alkaline phosphatase conjugate; Sigma) as above at 1:2500 dilution.
121 Membranes were washed for 1h in TBS-T, before developing at room temperature with
122 SigmaFAST BCIP/NBT prepared according to manufacturer's instructions (Sigma).

123 ***Animal Ethics.*** Animal work was approved by Griffith University Animal Ethics Committee
124 Protocol Number GLY/16/19/AEC. Animals were cared for and handled in accordance with
125 the guidelines of the Australian National Health and Medical Research Council (NHMRC).

126 ***Adherence and invasion assays with host cells.*** NTHi adherence and invasion was assessed
127 as previously (31, 32). Approximately 2.5 x 10^5 A549 cells were seeded into each well of a
128 flat-bottomed 24 well plate (Greiner, Germany) and allowed to settle overnight (37°C) before
129 inoculating with NTHi at an MOI of 30:1, or $8x10^6$ CFU in 250µL of RPMI media (Dubco)
130 containing 10% (v/v) foetal calf serum (FCS). Plates were incubated for 4h at 37°C with 5%
131 (v/v) CO_2. Wells were washed of non-adherent NTHi via multiple, gentle, washes with 1mL

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

132 of phosphate buffered saline (PBS). Visual checks were performed to ensure A549s were
133 intact, and planktonic NTHi were removed. Wells were then treated with 250µL of 0.25%
134 Trypsin-EDTA to dislodge adherent bacteria (5min 37°C) before serial dilution and drop
135 plating on columbia-blood agar (CBA) plates to enumerate bacterial loads. Results represent
136 triplicate values of biological duplicates. The percentage adherence was calculated from the
137 CFU in the inoculum.

138 Invasion assays were identical to the adherence assay with the following extra steps
139 following removal of adhered bacteria: extracellular bacteria were killed via treatment with
140 100µg/mL gentamicin in RPMI containing 10% (v/v) FCS for 1h at 37°C. Effectiveness of
141 gentamicin treatment was assessed by plating supernatant following treatment, with no
142 bacterial growth evident. Wells were then treated with 250µL 0.2% (v/v) Saponin to lyse
143 A549s (releasing intracellular bacteria). Visual checks were made to confirm cell lysis.
144 Surviving intracellular NTHi were enumerated via serial dilution and drop plating as per
145 adherence assays. Results represent triplicate values of biological duplicates. The percentage
146 invasion was calculated from the CFU in the inoculum.

147 ***Settling Assay.*** NTHi were grown in sBHI to an OD_{600} of 1.0. 3mL of OD_{600} 1.0 cells were
148 resuspended in PBS, mixed thoroughly and then split into triplicate cuvettes per variant.
149 Samples were monitored for 4 hours by measuring OD_{600}. Values were expressed as % of
150 initial reading.

151 ***Adherence assays with ECM components***

152 Flat bottom 96-well tissue culture treated plates (Falcon) were coated with vitronectin,
153 laminin or fibronectin (all Sigma-Aldrich), according to manufacturer protocols. Briefly,
154 working solutions of vitronectin (15 µg/ml) and laminin (6 µg/ml) were prepared in 1X
155 DPBS, whereas fibronectin was reconstituted in water (15 µg/ml). From the working stock of

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

156 vitronectin, 100 µl was added per well of 96 well plate and incubated at 37ºC for 2 hours
157 followed by overnight storage at 4ºC. For laminin and fibronectin, 100 µl of working
158 solutions was added to each well on the day of the assay, followed by immediate removal of
159 the solution. Wells were air dried for 45 min and washed twice with 1X DPBS prior to the
160 assay. Bacterial inoculum was prepared from log phase cultures of NTHi grown in sBHI and
161 added at a density of $5x10^6$ CFU/well prepared in 1X DPBS to wells coated with individual
162 ECM component. After incubation at 37ºC and 5% CO_2 for 1 hour, the supernatant was
163 removed, and wells were washed 4 times with 1X DPBS to remove any non-adherent
164 bacteria. Adherent bacteria were collected in 100 µl 1X DPBS with vigorous pipetting and
165 scraping of the wells. Dilutions of the collected sample as well as the inoculum were plated
166 on chocolate agar. The percentage adherence was calculated from the CFU in the inoculum.

167 ***Biofilm imaging and analysis***

168 Biofilms were formed by NTHi cultured within chambers of eight-well-chambered
169 coverglass slides (Thermo Scientific, Waltham, MA) as described previously (33). Briefly,
170 biofilms were formed by NTHi cultured within chambers of eight-well-chambered coverglass
171 slides (Thermo Scientific, Waltham, MA) using mid-log-phase NTHi cultures. Bacteria were
172 inoculated at 4×10^4 CFU in 200-µl final volume per well and incubated at 37°C with 5%
173 CO_2 for 24 h, with the growth medium replaced after 16 h. To visualize, biofilms were
174 stained with LIVE/DEAD BacLight stain (Life Technologies) and fixed overnight in fixative
175 (1.6% paraformaldehyde, 2.5% glutaraldehyde, and 4% acetic acid in 0.1 M phosphate buffer,
176 pH 7.4). Fixative was replaced with saline before imaging with a Zeiss 980 Meta-laser
177 scanning confocal microscope. Images were rendered with Zeiss Zen software. Z-stack
178 images were analyzed by COMSTAT2 (34) to determine biomass (µm3/µm2), average
179 thickness (µm), and roughness (Ra).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

180 ***Phylogenetic tree.*** The 16s rRNA sequence of fully annotated *H. influenzae* genomes
181 available in NCBI GenBank were aligned using CLUSTAL OMEGA (1.2.4).
182 **Supplementary Table 1** contains full details of strains, genes and data used.

183 ***Statistical Analysis.*** Graphs and statistics were generated via GraphPad Prism 5.0 (GraphPad
184 Software, La Jolla, California). Error bars represent standard deviation from mean values. A
185 one-way ANOVA was used to compare samples: p values of <0.05 (considered significant)
186 represented by *, p value of <0.001 indicated by **, p value of <0.001 indicated by ***.
187 Groups were considered not significantly different if $p > 0.05$ (no *).

188

189 **Results**

190 ***Lav expression is phase-variable in NTHi due to changes in length an SSR tract in the***
191 ***open-reading frame***

192 In order to study Lav function during colonisation and disease, we used prototype NTHi
193 strain 86-028NP (35) that encoded *lav* (NTHI0585) and enriched populations of bacteria via
194 single colony screening using fluorescent PCR (**Figure 1A**) for $GCAA_{(n)}$ SSR tract lengths
195 corresponding to all three possible reading frames. This resulted in three isogenic populations
196 enriched for tracts containing 21 GCAA repeats (21r), 22 GCAA repeats (22r), and 23 GCAA
197 repeats (23r) (**Figure 1B**). Analysis of *lav in silico* using the genome annotation from strain
198 86-028NP (Genbank accession number CP000057) determined that *lav* containing 21
199 $GCAA_{(n)}$ repeats would be in-frame and ON (expressed), and those populations where the
200 $GCAA_{(n)}$ tract was 22 or 23 repeats would be out-of-frame and OFF (not expressed), due to
201 premature transcriptional termination at stop codons in these two alternate reading frames.
202 We also cloned and over-expressed the predicted passenger domain of Lav from 86-028NP
203 (Lav-bind protein), based on previous analysis (22) to raise antisera. Western blots using this

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

204 antisera and the three enriched populations confirmed our prediction that 21 repeats was ON
205 (21r *lav* ON), and that 22 and 23 repeats were OFF (22r *lav* OFF and 23r *lav* OFF,
206 respectively), as we could only detect the Lav protein in the 21r population (**Figure 1C**).

207 ***The lav gene is switched OFF in NTHi isolates during both colonisation and invasive***
208 ***infection.***

209 We previously examined two collections of NTHi isolates for the expression of multiple
210 lipooligosaccharide (LOS) biosynthetic enzymes (36), demonstrating that certain enzymes
211 were selected for during invasive disease. We sought to further utilize these two collections
212 to determine if phase variation of the *lav* gene occurred during colonisation and invasive
213 disease. These two collections comprised carriage isolates, the ORChID collection (26, 27,
214 37), and a collection of invasive NTHi isolates (28). Fluorescently labelled PCR of the
215 $GCAA_{(n)}$ repeat tract of the *lav* gene (**Figure 1A**) was used to determine the ratio of each
216 tract length present in the bacterial population, and calculate the percentage ON/OFF ratio of
217 that population. Analysis of 16 isolates from our carriage collection showed that *lav* was
218 present in all strains, and predominantly OFF in these isolates (14/16; 87.5%) (**Table 1**).
219 Analysis of our invasive collection determined that the *lav* gene was present in ~69% of the
220 strains (**Table 1**), and where present, was also predominantly OFF (present in 50/72 isolates;
221 of the 50 isolates encoding a *lav* gene, the gene is OFF in 49/50; 98%). This indicates that
222 expression of *lav* may not be required during either colonisation or invasive infection, or
223 there is a direct selection against expression of the Lav protein during both phenotypic states.

224 ***Lav expression results in increased host cell adherence, but not invasion.***

225 In order to determine if Lav expression was required for an aspect of NTHi induced disease
226 other than nasopharyngeal colonisation or invasive infection, we investigated the broad role
227 of Lav during adherence to and invasion of the A549 human cell line, isolated from the lower

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

228 human airway, using our ON/OFF enriched populations. These assays demonstrated that the
229 Lav protein has a role in adherence to host cells, as 21r *lav* ON showed a significantly greater
230 percentage of adherence than both 22r and 23r *lav* OFF variants (**Figure 2A**, $p < 0.05$).
231 However, there was no significant difference in the ability of ON and OFF variants to invade
232 these same cells (**Figure 2A**). The CFU/well and MOI values for both adherence and
233 invasion assays are presented in **Supplementary Figure 2**. We also found that Lav
234 expression is not required for inter-bacterial adherence, as there was no difference in the rate
235 of settling as determined by using the optical density of a static culture of each of our
236 enriched variants over 4 hours (**Figure 2B**).

237 ***Lav is not required for adherence to ECM components***

238 Epithelial cells of the human respiratory tract produce multiple extracellular matrix (ECM)
239 components (38-40). Since *lav* ON/OFF status affected the ability of NTHi to adhere to lung
240 epithelial cells (**Figure 2A**), we tested adherence of the *lav* variants to the ECM components
241 laminin, fibronectin and vitronectin for 1 hour. There was no significant difference observed
242 between the percent adherence of the variants to laminin, fibronectin or vitronectin (**Figure**
243 **2C**), indicating that Lav is not involved in adherence of NTHi to these ECM components, but
244 may instead be required to adhere specifically to receptor(s) only present on host cells.

245 ***Lav expression results in biofilms with greater biomass and thickness***

246 To determine if Lav phase variation resulted in differences in biofilm formation, a key feature
247 of NTHi pathology, biofilms of our enriched 21r *lav* ON, 22r *lav* OFF, and 23r *lav* OFF
248 variants were grown for 24 hours. Biofilms formed by the 21r *lav* ON variant exhibited
249 significantly greater biomass and average thickness compared to variants that did not express
250 Lav (22r and 23r *lav* OFF; **Figure 3A, 3B**). Biofilms formed by 22r *lav* OFF tended to have
251 an architecture that was rougher in comparison to either of the other variants, likely due to the

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

252 more dispersed nature of these biofilms, but the roughness of all three variants was
253 statistically similar (**Figure 3C**). Based on gross biofilm abundance and microscopic analysis
254 (**Figure 3D**), NTHi that expressed Lav (21r *lav* ON) formed significantly larger biofilms
255 overall.

256 ***lav* distribution and conservation in *Haemophilus spp*.**

257 Previous studies demonstrated a broad distribution of *lav*, and multiple allelic variants of the
258 Lav passenger domain in *Haemophilus* spp. (21, 22). However, there was no consistent
259 naming of these variants in *Haemophilus* spp., nor a thorough analysis of the distribution or
260 variability present. Therefore, we examined all fully annotated *Haemophilus* spp. genomes
261 available in NCBI GenBank. There were 73 fully annotated *H. influenzae* genomes available
262 at the time of this investigation. Of those 73, 47 were NTHi and the remainder were either
263 typeable (serotypes a-f) or the serotype was undetermined. The *lav* gene was present in 29/47
264 NTHi genomes (~62% gene presence), very similar to that observed in our invasive
265 collection (69% presence). Interestingly, *lav* was absent in all strains annotated as serotype b-
266 f, but was present in all strains (4) annotated as *H. influenzae* serotype a. A *lav* homologue,
267 named *las*, was present in all 11 available genomes of *H. influenzae* biogroup *aegyptius* (7
268 fully annotated, plus 4 available genomes) (**Figure 4**).

269 Further, we carried out detailed sequence analysis of all *lav* genes from these 73
270 *Haemophilus* spp. genomes, as previous work in *Neisseria* spp. had identified a number of
271 allelic variants (21). Passenger domain variants 1 and 2 (previously named AutB1 and AutB2
272 (21)) were found exclusively in strains annotated as NTHi, with an approximate 60:40 split
273 (59% and 41%, respectively). There appears to be a lineage distribution of these variants,
274 with closely related strains containing the same passenger domain allele (**Figure 4**).
275 Alignment of the sequences of the Lav passenger domain (**Figure 5**) showed that they were

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

276 more diverse than previously described (21). Analysis of variant 1 showed two sub-variants
277 present, with only 73.06% identity, and which we propose to name variants 1.1 and 1.2
278 (alignment in **Supplementary Figure 3**). With the exception of one *H. haemolyticus* strain,
279 Lav passenger domain variant 3 (previously named AutB3 (21)) was found exclusively in *H.*
280 *influenzae* biogroup *aegyptius* strains. Our sequence analysis of variant 3 also showed two
281 distinct sub-variants, showing only 51.26% identity, and therefore we have also proposed
282 delineating variant 3 into variants 3.1 and 3.2 (alignment in **Supplementary Figure 3**).

283

284 **Discussion**

285 Surface exposed NTHi phase-variable autotransporters are important virulence determinants
286 (41). A Lav homologue, named AutB, was shown to be highly diverse, encoding multiple
287 allelic variants of the functional passenger domain (21), with AutB important for biofilm
288 formation in *N. meningitidis*. Previous work demonstrated that the *lav* gene in NTHi phase-
289 varied OFF during repeated infection (22). Therefore, we aimed to determine the phenotypic
290 role of Lav in NTHi, and to rationalise the prevalence and diversity of Lav in *Haemophilus*
291 spp.

292 Analysis of our carriage and invasive NTHi collections (36) revealed *lav* to be present in
293 ~69% (50/72) of strains in our invasive collection (**Table 1A**), but in every strain in our
294 carriage collection (**Table 1B**). Therefore, it appears that the invasive collection is
295 representative of all NTHi strains, with *lav* found in ~62% of fully annotated NTHi genomes
296 (**Figure 4**). The high proportion of Lav observed in our carriage collection is likely an
297 artefact of a small (16 isolates) sample size. Analysis of our invasive and carriage collections
298 showed that *lav* is predominantly phase-varied OFF in NTHi colonising the nasopharynx
299 (carriage collection; 87.5%) and during invasive infection (where present, *lav* is OFF in 49/50

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

300 isolates, equating to 98%), suggesting that Lav is either not required or directly selected
301 against during both colonisation and invasive infection, or expressed in a distinct niche. Our
302 invasive collection also represents just a 'snap-shot' of the exact phenotypic state at a
303 particular time point during invasive infection, i.e., when treatment is required, which is
304 likely to be during the later stages of disease. Previous work has shown that the *lav* gene
305 switches OFF over subsequent episodes of infection (25), but can rapidly change expression
306 over short periods (24). As we have found it to be OFF in the majority of both carriage and
307 invasive isolates, it is possible that selection for the OFF state is due to negative selection
308 from immune detection/pressure. Immune selection against multiple outer-membrane
309 proteins has been reported in *N. meningitidis* (42), with gene expression phase-varying from
310 ON to OFF during persistent carriage. Similar work with an additional phase-variable
311 autotransporter in NTHi, Hia, showed Hia phase-varies OFF during opsonophagocytic killing
312 (17). Our phenotypic findings regarding biofilm formation are also in agreement with
313 previous work involving the Lav homologue AutB in *Neisseria meningitidis* (21). It therefore
314 appears that Lav/AutB play a similar role in both species. The establishment of bacterial
315 biofilms is critical during colonization and disease. Biofilms help bacteria adhere to the
316 mucosal surfaces, and provide increased resistance to host defences and antimicrobials (43-
317 45). Thus, expression of Lav might provide a selective advantage to NTHi during initial
318 colonisation and establishment at the mucosal surface. Once established, factors such as
319 immune pressures or microenvironmental conditions may select against the Lav expressing
320 subpopulation as observed in the persistently colonized or invasive isolates assessed.

321 Our phenotypic analysis demonstrated that Lav has a role in adherence to, but not for
322 invasion of, human A549 lung cells (**Figure 2A**), and that Lav expression does not play a role
323 in adherence to ECM proteins (**Figure 2C**). This indicates that there is a state in which Lav
324 expression is beneficial, perhaps during the early stages of colonisation or during progression

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

325 from the upper to the lower respiratory tract, or whilst establishing invasive/systemic
326 infection. Another possibility is that Lav interacts with a specific receptor present on human
327 cells, and does not directly bind to the ECM proteins fibronectin, laminin, and vitronectin.
328 The NTHi autotransporter Hia has been shown to bind to human specific glycans rather than
329 proteins as high affinity receptors (46).

330 Understanding the prevalence and conservation of Lav is key for determining its suitability
331 for use in a rationally designed subunit vaccine against NTHi. Our analysis determined that
332 the *lav* gene is found in ~62% of *Haemophilus* spp. Intriguingly, there was no *lav* gene, or
333 close homologue, in any strain annotated as *H. influenzae* serotype b-f. The presence of Lav
334 in *H. influenzae* serotype a only may suggest that this protein is an important virulence factor
335 in these strains, although the small number of sequences publicly available for analysis means
336 this hypothesis will require further investigation.

337 Our investigation of the diversity of the Lav passenger domains present in *Haemophilus* spp.
338 showed that there are five different allelic variants of the Lav passenger domain (the
339 functional extracellular region) present in *Haemophilus* spp. Our detailed sequence analysis
340 showed that both variants 1 and 3 can be further divided into two separate allelic variants –
341 1.1 and 1.2, and 3.1 and 3.2 (**Figure 5**). Variants 3.1 and 3.2, previously annotated as Las
342 (22, 23), are found exclusively in *H. influenzae* biogroup *aegyptius* isolates.

343 *H. influenzae* biogroup *aegyptius* strains cause the invasive disease Brazilian purpuric fever
344 (BPF), a meningitis-like disease with high fatality (47). The ubiquitous presence of Lav
345 variants 3.1 and 3.2 in *H. influenzae* biogroup *aegyptius* isolates supports the idea that these
346 particular variants contribute to the development of BPF, although it has previously been
347 reported that no single factor is required for BPF (47). It has also previously been reported
348 that *las* expression is highly variable during an animal model of BPF (24), with expression of

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

349 the gene shown to decrease (switch OFF) after 24 hours, then increase (switch ON) at the 48
350 hour time point post infection, demonstrating complex regulation of Lav/Las occurs during
351 disease. In summary, our analysis has shown that there are five unique variants of the Lav
352 passenger domain encoded by *Haemophilus* spp., and there is a distinct distribution between
353 serotype/species. Future investigation into the functional differences between passenger
354 domain variants is needed to determine if these variants have different functions.

355 It is important to understand the role of bacterial surface factors like Lav in order to
356 understand NTHi-mediated diseases, and to develop effective vaccines and treatments. Our
357 work has determined that expression of a particular Lav variant (1.2 in strain 86-028NP)
358 results in greater host cell adherence and biofilm formation, and demonstrated that the *lav*
359 gene is present as five different allelic variants in *Haemophilus* spp. As Lav is present in
360 ~60% of NTHi strains, understanding the role of all variants is key to understanding NTHi
361 disease, and further work is required to assess if Lav can form part of a multi-subunit,
362 rationally designed vaccine against NTHi.

363

364 **Acknowledgements**

365 We thank the Department of Cell Biology, Neurobiology and Anatomy at MCW for the use
366 of their Zeiss LSM980 confocal microscope. This work was supported by an Australian
367 Research Council (ARC) Discovery Project grant (DP180100976) to J.M.A. an Australian
368 National Health and Medical Research Council (NHMRC) Principal Research Fellowship
369 (1138466) to M.P.J, and a National Institutes of Health (NIH) grant (R21-DC016709) to
370 K.L.B. The ORChID study was supported by a NHMRC project grant (GNT615700) and a
371 program grant from the Children's Health Foundation Queensland (5006). Publication costs

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

372 of this work were supported by a generous donation from the Bourne Foundation, Melbourne,
373 Australia.

374

375

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

376 **References**

377 1. Murphy TF, Faden H, Bakaletz LO, Kyd JM, Forsgren A, Campos J, Virji M, Pelton
378 SI. 2009. Nontypeable *Haemophilus influenzae* as a pathogen in children. Pediatr
379 Infect Dis J 28:43-8.
380 2. Sethi S, Murphy TF. 2008. Infection in the pathogenesis and course of chronic
381 obstructive pulmonary disease. N Engl J Med 359:2355-65.
382 3. Van Eldere J, Slack MP, Ladhani S, Cripps AW. 2014. Non-typeable *Haemophilus*
383 *influenzae*, an under-recognised pathogen. Lancet Infect Dis 14:1281-92.
384 4. Johnson RH. 1988. Community-acquired pneumonia: etiology, diagnosis, and
385 treatment. Clin Ther 10:568-73.
386 5. Ladhani S, Slack MP, Heath PT, von Gottberg A, Chandra M, Ramsay ME. 2010.
387 Invasive Haemophilus influenzae Disease, Europe, 1996-2006. Emerg Infect Dis
388 16:455-63.
389 6. Collins S, Vickers A, Ladhani SN, Flynn S, Platt S, Ramsay ME, Litt DJ, Slack MP.
390 2016. Clinical and Molecular Epidemiology of Childhood Invasive Nontypeable
391 Haemophilus influenzae Disease in England and Wales. Pediatr Infect Dis J 35:e76-
392 84.
393 7. Slack MPE, Cripps AW, Grimwood K, Mackenzie GA, Ulanova M. 2021. Invasive
394 Haemophilus influenzae Infections after 3 Decades of Hib Protein Conjugate Vaccine
395 Use. Clin Microbiol Rev 34:e0002821.
396 8. Atkinson CT, Kunde DA, Tristram SG. 2017. Expression of acquired macrolide
397 resistance genes in Haemophilus influenzae. J Antimicrob Chemother 72:3298-3301.
398 9. Phillips ZN, Tram G, Seib KL, Atack JM. 2019. Phase-variable bacterial loci: how
399 bacteria gamble to maximise fitness in changing environments. Biochem Soc Trans
400 47:1131-1141.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

401 10. Phillips ZN, Husna AU, Jennings MP, Seib KL, Atack JM. 2019. Phasevarions of
402 bacterial pathogens - phase-variable epigenetic regulators evolving from restriction-
403 modification systems. Microbiology doi:10.1099/mic.0.000805.
404 11. Moxon R, Bayliss C, Hood D. 2006. Bacterial contingency loci: the role of simple
405 sequence DNA repeats in bacterial adaptation. Annu Rev Genet 40:307-33.
406 12. Jeffrey N. Weuser DJM, Peter D. Butler, Alf A. Lindberg, E. Richard Moxon. 1990.
407 Characterization of repetitive sequences controlling phase variation of Haemophilus
408 influenzae Lipopolysaccharide. Journal of Bacteriology:3304 - 3309.
409 13. Elango D, Schulz BL. 2020. Phase-Variable Glycosylation in Nontypeable
410 Haemophilus influenzae. J Proteome Res 19:464-476.
411 14. Meuskens I, Saragliadis A, Leo JC, Linke D. 2019. Type V Secretion Systems: An
412 Overview of Passenger Domain Functions. Frontiers in Microbiology 10.
413 15. Battaglioli EJ, Goh KGK, Atruktsang TS, Schwartz K, Schembri MA, Welch RA.
414 2018. Identification and Characterization of a Phase-Variable Element That Regulates
415 the Autotransporter UpaE in Uropathogenic Escherichia coli. mBio 9.
416 16. Spahich NA, Hood DW, Moxon ER, St Geme JW, 3rd. 2012. Inactivation of
417 Haemophilus influenzae lipopolysaccharide biosynthesis genes interferes with outer
418 membrane localization of the hap autotransporter. J Bacteriol 194:1815-22.
419 17. Atack JM, Winter LE, Jurcisek JA, Bakaletz LO, Barenkamp SJ, Jennings MP. 2015.
420 Selection and counter-selection of Hia expression reveals a key role for phase-
421 variable expression of this adhesin in infection caused by non-typeable *Haemophilus*
422 *influenzae*. J Infect Dis 212:645-53.
423 18. Serruto D, Spadafina T, Ciucchi L, Lewis LA, Ram S, Tontini M, Santini L, Biolchi
424 A, Seib KL, Giuliani MM, Donnelly JJ, Berti F, Savino S, Scarselli M, Costantino P,
425 Kroll JS, O'Dwyer C, Qiu J, Plaut AG, Moxon R, Rappuoli R, Pizza M, Aricò B.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

426 2010. Neisseria meningitidis GNA2132, a heparin-binding protein that induces
427 protective immunity in humans. Proceedings of the National Academy of Sciences of
428 the United States of America 107:3770-3775.
429 19. van Ulsen P, van Alphen L, ten Hove J, Fransen F, van der Ley P, Tommassen J.
430 2003. A Neisserial autotransporter NalP modulating the processing of other
431 autotransporters. Mol Microbiol 50:1017-30.
432 20. Arenas J, Cano S, Nijland R, van Dongen V, Rutten L, van der Ende A, Tommassen
433 J. 2015. The meningococcal autotransporter AutA is implicated in autoaggregation
434 and biofilm formation. Environ Microbiol 17:1321-37.
435 21. Arenas J, Paganelli FL, Rodríguez-Castaño P, Cano-Crespo S, van der Ende A, van
436 Putten JPM, Tommassen J. 2016. Expression of the Gene for Autotransporter AutB of
437 Neisseria meningitidis Affects Biofilm Formation and Epithelial Transmigration.
438 Frontiers in Cellular and Infection Microbiology 6.
439 22. Davis J SA, Hughes WR, Golomb M. . 2001. Evolution of an autotransporter: domain
440 shuffling and lateral transfer from pathogenic Haemophilus to Neisseria. J Bacteriol
441 2001;183(15):4626–4635. .
442 23. Cury GCG, Pereira RFC, de Hollanda LM, Lancellotti M. 2015. Inflammatory
443 response of Haemophilus influenzae biotype aegyptius causing Brazilian Purpuric
444 Fever. Brazilian journal of microbiology : [publication of the Brazilian Society for
445 Microbiology] 45:1449-1454.
446 24. Pereira RFC, Theizen TH, Machado D, Guarnieri JPO, Gomide GP, Hollanda LM,
447 Lancellotti M. 2020. Analysis of potential virulence genes and competence to
448 transformation in Haemophilus influenzae biotype aegyptius associated with Brazilian
449 Purpuric Fever. Genet Mol Biol 44:e20200029.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

450 25. Harrison A, Hardison RL, Fullen AR, Wallace RM, Gordon DM, White P, Jennings
451 RN, Justice SS, Mason KM. 2020. Continuous Microevolution Accelerates Disease
452 Progression during Sequential Episodes of Infection. Cell Rep 30:2978-2988.e3.
453 26. Lambert SB, Ware RS, Cook AL, Maguire FA, Whiley DM, Bialasiewicz S, Mackay
454 IM, Wang D, Sloots TP, Nissen MD, Grimwood K. 2012. Observational Research in
455 Childhood Infectious Diseases (ORChID): a dynamic birth cohort study. BMJ Open
456 2.
457 27. Sarna M, Lambert SB, Sloots TP, Whiley DM, Alsaleh A, Mhango L, Bialasiewicz S,
458 Wang D, Nissen MD, Grimwood K, Ware RS. 2018. Viruses causing lower
459 respiratory symptoms in young children: findings from the ORChID birth cohort.
460 Thorax 73:969-979.
461 28. Staples M, Graham RMA, Jennison AV. 2017. Characterisation of invasive clinical
462 Haemophilus influenzae isolates in Queensland, Australia using whole-genome
463 sequencing. Epidemiol Infect 145:1727-1736.
464 29. Phillips ZN, Brizuela C, Jennison AV, Staples M, Grimwood K, Seib KL, Jennings
465 MP, Atack JM. 2019. Analysis of invasive non-typeable *Haemophilus influenzae*
466 isolates reveals a selection for the expression state of particular phase-variable
467 lipooligosaccharide biosynthetic genes. Infect Immun 87:10.1128/iai.00093-19.
468 30. Tram G, Jen FE, Phillips ZN, Timms J, Husna AU, Jennings MP, Blackall PJ, Atack
469 JM. 2021. Streptococcus suis Encodes Multiple Allelic Variants of a Phase-Variable
470 Type III DNA Methyltransferase, ModS, That Control Distinct Phasevarions.
471 mSphere 6:e00069-21.
472 31. Goyal M, Singh M, Ray P, Srinivasan R, Chakraborti A. 2015. Cellular interaction of
473 nontypeable Haemophilus influenzae triggers cytotoxicity of infected type II alveolar
474 cells via apoptosis. Pathogens and Disease 73:1-12.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

475 32. Geme JWS, Falkow S. 1990. Haemophilus influenzae adheres to and enters cultured
476 human epithelial cells. Infection and Immunity 58:4036-4044.
477 33. Jurcisek JA, Dickson AC, Bruggeman ME, Bakaletz LO. 2011. *In vitro* biofilm
478 formation in an 8-well chamber slide. Journal of Visualized Experiments
479 doi:10.3791/2481:pii: 2481. doi: 10.3791/2481.
480 34. Heydorn A, Nielsen AT, Hentzer M, Sternberg C, Givskov M, Ersbøll BK, Molin S.
481 2000. Quantification of biofilm structures by the novel computer program
482 COMSTAT. Microbiology (Reading) 146 (Pt 10):2395-2407.
483 35. Harrison A, Dyer DW, Gillaspy A, Ray WC, Mungur R, Carson MB, Zhong H,
484 Gipson J, Gipson M, Johnson LS, Lewis L, Bakaletz LO, Munson RS, Jr. 2005.
485 Genomic sequence of an otitis media isolate of nontypeable Haemophilus influenzae:
486 comparative study with H. influenzae serotype d, strain KW20. J Bacteriol 187:4627-
487 36.
488 36. Phillips ZN, Brizuela C, Jennison AV, Staples M, Grimwood K, Seib KL, Jennings
489 MP, Atack JM. 2019. Analysis of Invasive Nontypeable Haemophilus influenzae
490 Isolates Reveals Selection for the Expression State of Particular Phase-Variable
491 Lipooligosaccharide Biosynthetic Genes. Infect Immun 87.
492 37. Palmu AA, Ware RS, Lambert SB, Sarna M, Bialasiewicz S, Seib KL, Atack JM,
493 Nissen MD, Grimwood K. 2019. Nasal swab bacteriology by PCR during the first 24-
494 months of life: A prospective birth cohort study. Pediatr Pulmonol 54:289-296.
495 38. Coraux C, Roux J, Jolly T, Birembaut P. 2008. Epithelial cell-extracellular matrix
496 interactions and stem cells in airway epithelial regeneration. Proc Am Thorac Soc
497 5:689-94.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

498 39. Peters-Hall JR, Brown KJ, Pillai DK, Tomney A, Garvin LM, Wu X, Rose MC. 2015.
499 Quantitative proteomics reveals an altered cystic fibrosis in vitro bronchial epithelial
500 secretome. Am J Respir Cell Mol Biol 53:22-32.
501 40. Salazar-Peláez LM, Abraham T, Herrera AM, Correa MA, Ortega JE, Paré PD, Seow
502 CY. 2015. Vitronectin expression in the airways of subjects with asthma and chronic
503 obstructive pulmonary disease. PLoS One 10:e0119717.
504 41. Spahich NA, St Geme JW, 3rd. 2011. Structure and function of the Haemophilus
505 influenzae autotransporters. Frontiers in cellular and infection microbiology 1:5-5.
506 42. Alamro M, Bidmos FA, Chan H, Oldfield NJ, Newton E, Bai X, Aidley J, Care R,
507 Mattick C, Turner DPJ, Neal KR, Ala'Aldeen DAA, Feavers I, Borrow R, Bayliss
508 CD. 2014. Phase Variation Mediates Reductions in Expression of Surface Proteins
509 during Persistent Meningococcal Carriage. Infection and Immunity 82:2472-2484.
510 43. Yan J, Bassler BL. 2019. Surviving as a Community: Antibiotic Tolerance and
511 Persistence in Bacterial Biofilms. Cell Host Microbe 26:15-21.
512 44. Murphy TF, Bakaletz LO, Smeesters PR. 2009. Microbial interactions in the
513 respiratory tract. Pediatr Infect Dis J 28:S121-6.
514 45. Novotny LA, Brockman KL, Mokrzan EM, Jurcisek JA, Bakaletz LO. 2019. Biofilm
515 biology and vaccine strategies for otitis media due to nontypeable Haemophilus
516 influenzae. J Pediatr Infect Dis 14:69-77.
517 46. Atack JM, Day CJ, Poole J, Brockman KL, Timms JRL, Winter LE, Haselhorst T,
518 Bakaletz LO, Barenkamp SJ, Jennings MP. 2020. The Nontypeable Haemophilus
519 influenzae Major Adhesin Hia Is a Dual-Function Lectin That Binds to Human-
520 Specific Respiratory Tract Sialic Acid Glycan Receptors. mBio 11.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

521 47. Harrison LH, Simonsen V, Waldman EA. 2008. Emergence and disappearance of a
522 virulent clone of Haemophilus influenzae biogroup aegyptius, cause of Brazilian
523 purpuric fever. Clin Microbiol Rev 21:594-605.

524

525

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

526 **Table 1. Fragment Length Analysis of (A) Invasive and (B) Carriage Collections**
527 **screened for *lav* SSR tract length using Lav_F and Lav_R.** The results shown indicate
528 whether the *lav* gene was ON (>70% ON; green), OFF (>70% OFF; red), mixed ON and OFF
529 (orange), or if there was no gene (Grey) as we could not amplify a PCR product.

A) Invasive Collection				
OFF	**ON**	**Mixed**	**No gene**	**Total**
49	1	0	22	**72**
68.06	1.39	0.00	30.56	
Gene presence: 69.4% (50/72)				
B) Carriage Collection				
OFF	**ON**	**Mixed**	**No gene**	**Total**
14	2	0	0	**16**
87.5%	12.5%	0%	0%	
Gene presence: 100% (16/16)				

530
531
532
533

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

534 **Figure Legends**

535 **Figure 1. Phase variation of the *lav* gene. (A)**. The 2.2kb *lav* gene, with 5`GCAA$_{(n)}$ simple
536 sequence repeat (SSR) tract in grey. A fluorescently (FAM) labelled forward primer (Lav_F;
537 FAM indicated by green hexagon) binds upstream of the SSR tract. The reverse primer
538 (Lav_R) binds downstream of the SSR tract. (**B)** fragment analysis traces of enriched
539 variants for three consecutive GCAA$_{(n)}$ repeat tract lengths (21r, 22r, 23r). (**C**). Western Blot
540 using whole-cell lysates of 86-028NP isogenic strains enriched in (B) with the *lav* gene
541 containing an SSR number of 21, 22 or 23 repeats. An SSR tract number of 21 (21r) puts the
542 gene in-frame, and ON, indicated by presence of the Lav protein detected using anti-Lav
543 antisera. The 22r and 23r populations have the *lav* gene out-of-frame and OFF, with no Lav
544 detected in cell lysates of these strains.

545 **Figure 2. Impact of Lav expression on adherence and invasion.** Impact of Lav expression
546 on (**A**) adherence to and (**B**) invasion of the human A549 cell line was evaluated with Lav
547 ON/OFF variants in NTHi strain 86-028NP, which expresses the Lav 1.2 variant. (**C**) Auto-
548 aggregation was investigated by monitoring the OD_{600} of static cultures of our ON/OFF
549 variants. **(D)** Adherence of NTHi *lav* variants to ECM proteins Fibronectin, Laminin, and
550 Vitronectin. Statistical analysis was carried out by one way ANOVA. Error bars represent
551 standard deviation from mean values; p value of <0.05 represented by *; p value of <0.001
552 indicated by ***; NSD = no significant difference between any of strains.

553 **Figure 3. Impact of Lav expression on biofilm formation. (A)** Biomass, **(B)** average
554 thickness, and **(C)** roughness of biofilms grown for 24 h. Biofilms formed by the Lav
555 expressing variant were of significantly greater biomass and thickness than those formed by
556 the Lav non-expressing variants. Biofilms were analyzed by COMSTAT2, and values are
557 shown as mean ± standard error of the mean. Statistical analysis was carried out by one way

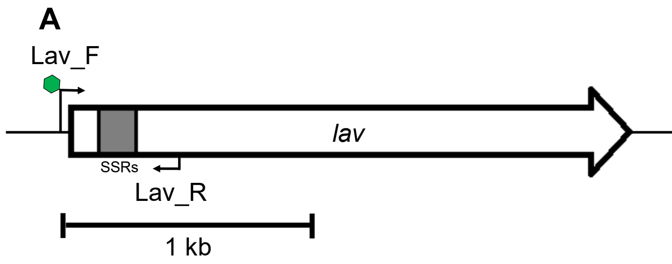
bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

558 ANOVA. Error bars represent standard deviation from mean values; p value of <0.05
559 (considered significant) represented by *, p value of <0.01 indicated by **; p value of <0.001
560 indicated by ***; NSD = no significant difference between any of the strains. **(D)**
561 Representative low-magnification images of biofilm density and distribution. Biofilms
562 formed by 21r *lav* ON appeared denser with more and larger tower like structures compared
563 to 22r *lav* OFF. 23r *lav* OFF formed biofilms with an intermediate distribution and smaller
564 tower like structures. Bacteria are shown in green. Scale bar = 500 μm.

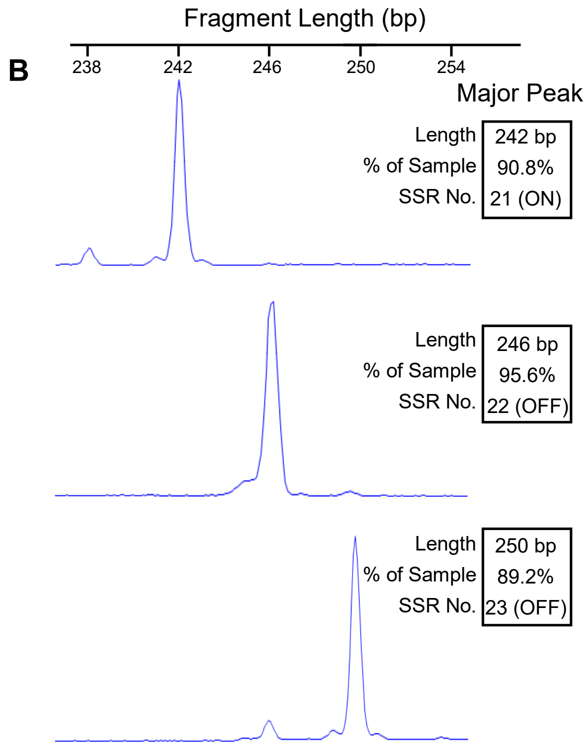
565 **Figure 4. Distribution of *lav* gene in *Haemophilus spp*.** Phylogenetic tree generated by
566 CLUSTAL OMEGA (1.2.4) using 16s rRNA gene sequences from NTHi and *H. influenzae*
567 serotype a GenBank entries. For sequences see **Supplementary Table 1**. Prefix describes
568 species: NTHi – Non-Typeable *Haemophilus influenzae*; Hia – *Haemophilus influenzae*
569 *serotype a*; aeg – *Haemophilus influenzae* biogroup *aegyptius*; Hh – *Haemophilus*
570 *haemolyticus*. Suffix indicates strain name. i.e. "NTHi_86028NP" is Non-Typeable
571 *Haemophilus influenzae* strain 86-028NP. Included in the figure is the Lav passenger domain
572 allele form (**1.1, 1.2, 2, 3.1, 3.2**) and a negative column (**-ve**) to show genomes that did not
573 contain the *lav* gene.

574 **Figure 5. Alignment of the Lav Passenger Domain Alleles.** CLUSTAL OMEGA (1.2.4)
575 was used to align the distinct Lav passenger domain allele forms using representative amino
576 acid sequences from five strains – R2866 (1.1), 86-028NP (1.2), 375 (2), HE40 (3.1) and
577 HE07 (3.2).

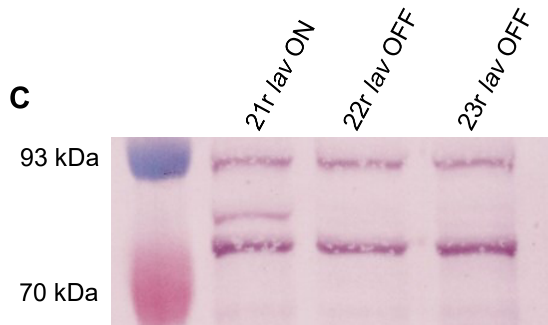
A



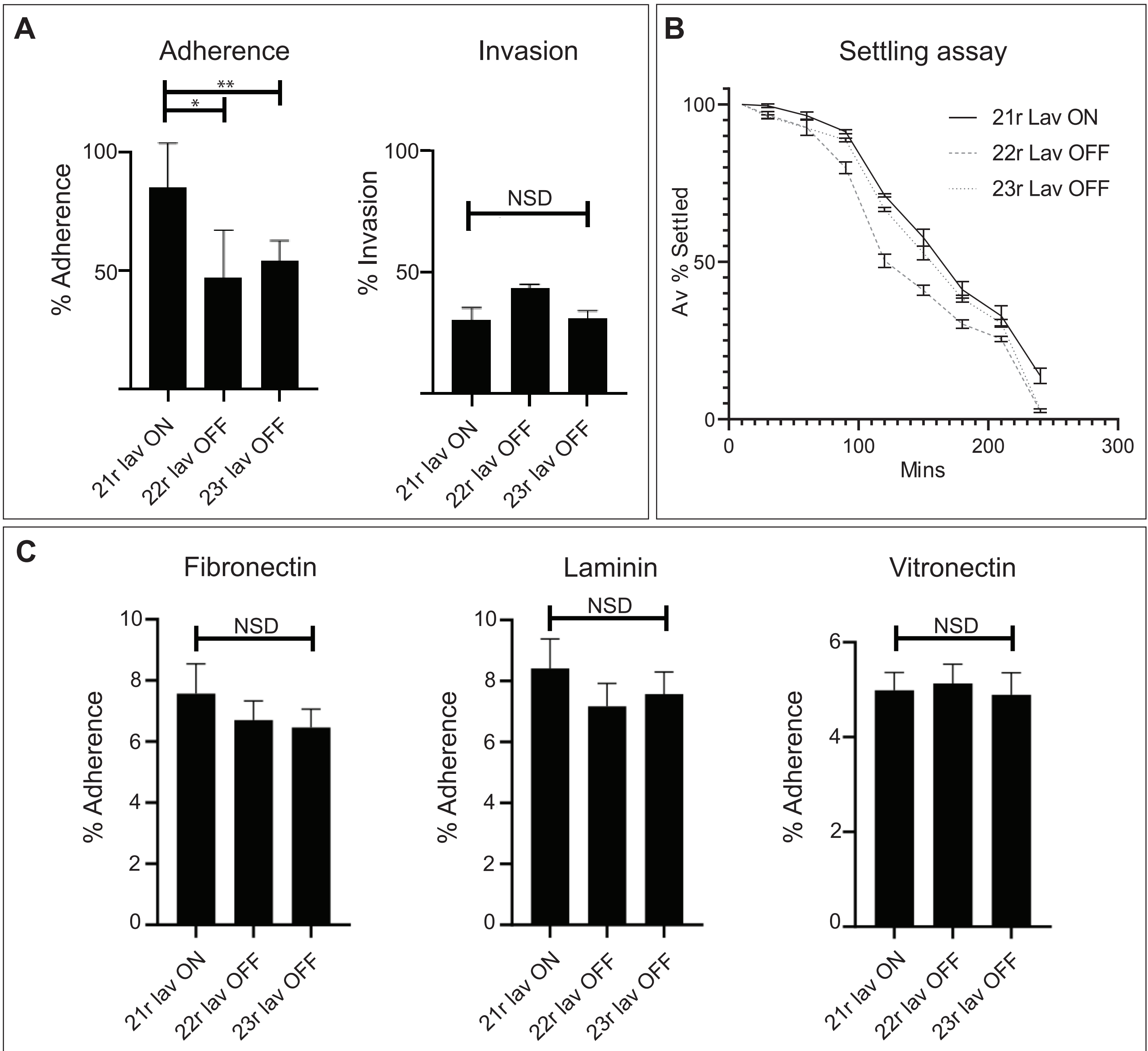
B

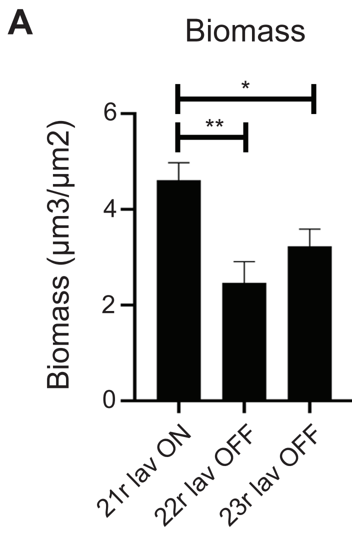


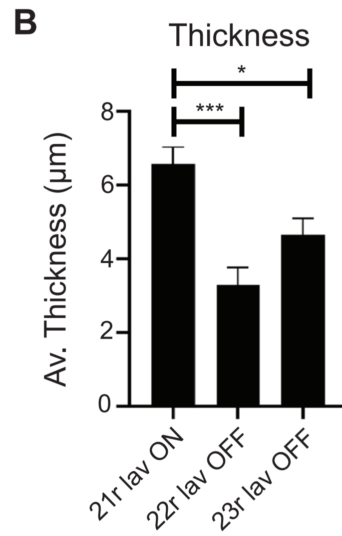
C

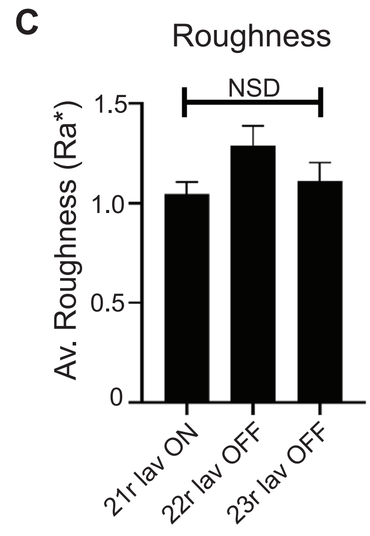


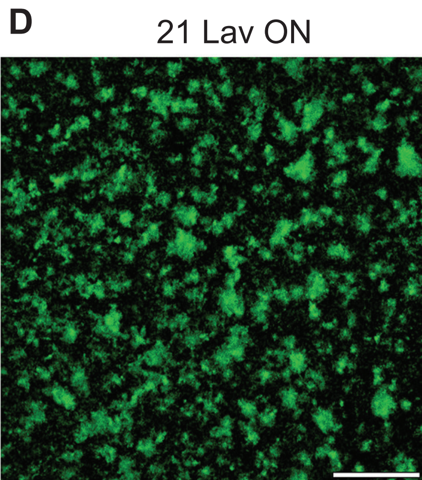
bioRxiv preprint doi: https://doi.org/10.1101/2021.10.15.464623; this version posted October 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

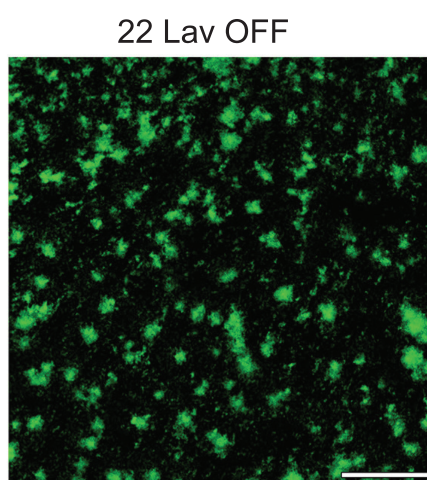


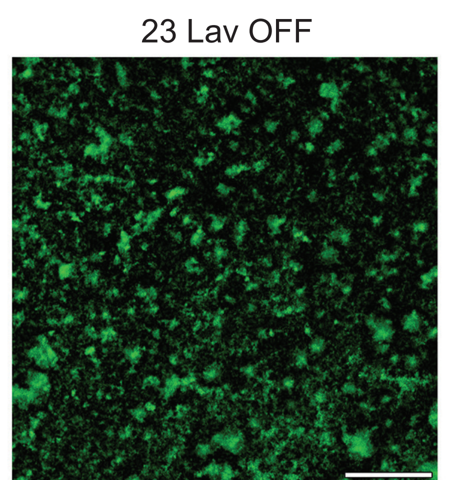
A


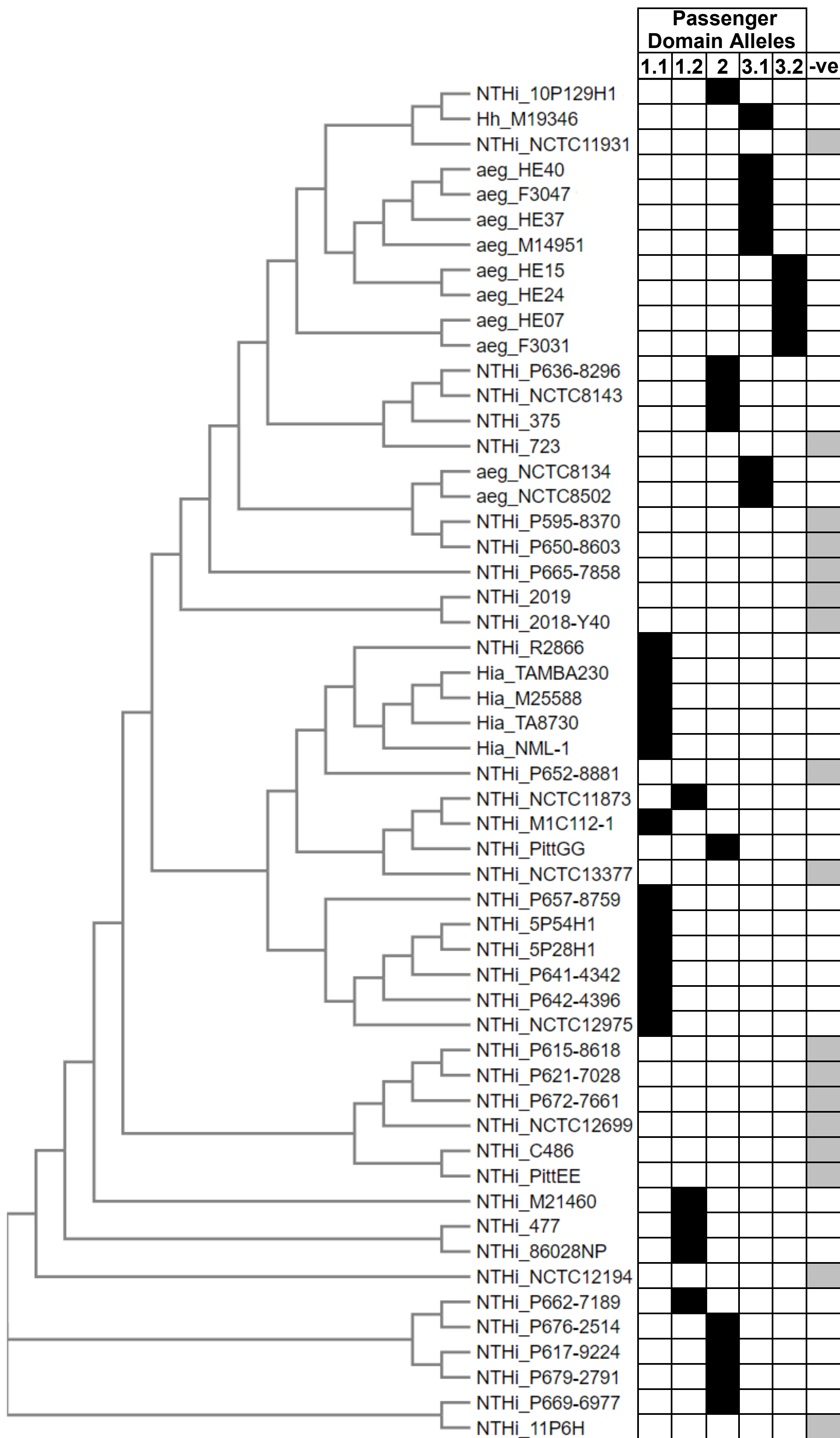
B


C


D







Passenger Domain Alleles
1.1
1.2
2
3.1
3.2
-ve
NTHi_10P129H1
Hh_M19346
NTHi_NCTC11931
aeg_HE40
aeg_F3047
aeg_HE37
aeg_M14951
aeg_HE15
aeg_HE24
aeg_HE07
aeg_F3031
NTHi_P636-8296
NTHi_NCTC8143
NTHi_375
NTHi_723
aeg_NCTC8134
aeg_NCTC8502
NTHi_P595-8370
NTHi_P650-8603
NTHi_P665-7858
NTHi_2019
NTHi_2018-Y40
NTHi_R2866
Hia_TAMBA230
Hia_M25588
Hia_TA8730
Hia_NML-1
NTHi_P652-8881
NTHi_NCTC11873
NTHi_M1C112-1
NTHi_PittGG
NTHi_NCTC13377
NTHi_P657-8759
NTHi_5P54H1
NTHi_5P28H1
NTHi_P641-4342
NTHi_P642-4396
NTHi_NCTC12975
NTHi_P615-8618
NTHi_P621-7028
NTHi_P672-7661
NTHi_NCTC12699
NTHi_C486
NTHi_PittEE
NTHi_M21460
NTHi_477
NTHi_86028NP
NTHi_NCTC12194
NTHi_P662-7189
NTHi_P676-2514
NTHi_P617-9224
NTHi_P679-2791
NTHi_P669-6977
NTHi_11P6H

```
              Passenger
           Domain Allele              SSR Tract
NTHi_R2866     1.1   ASKQASKQASKQASKQASKQKFKKSFILSLLFSALYSSPLLA-V-DYVYDKTKLT-----  53
NTHi_86028NP   1.2   ASKQASKQASKQASKQASKQKFKKSFIISLFFSILYTSPLLAVV-DYTYDNSKLT-----  54
NTHi_375        2    ASKQASKQASKQASKQASKQKFKKSFIVSLFFSILYTSPLLA-V-DYVYDKTKLT-----  53
aeg_HE40       3.1   ASKQASKQASKQASKQASKQKFSLIYSLLIAFLFIPYSV----SGEPKYVQEYFVKERGI  56
aeg_HE07       3.2   ASKQASKQASKQASKQASKQKFSLISILASTFLFISNTSFAAAGDVPAYITQYLTHEKKE  60
                     **********************.    :   *  :  :          *    :.

NTHi_R2866     1.1   --------DDEITRLKKLRDRDSEYWKEETYFIKSTPQSPPNIPR--------LFP-KNS  96
NTHi_86028NP   1.2   --------NEQIERLKKLRDRNSEYWKKETYLLKSPPSNFPDIPV--------LFP-KDS  97
NTHi_375        2    --------DDEITRLKKLRDRNSEYWKDDLFRLDIPKETGLRHDIKGAATGNFSYPVIGS  105
aeg_HE40       3.1   QYFGEGSGASVFYSLKDMLDRNSVAWSPVPKQII-------DYITNT-EACYFSHP----  104
aeg_HE07       3.2   Q-TGDYWHYYYTYSLKSMQNPDSIVWKPVPQKII-------DGLIKGWKACQN-------  105
                                 **.: : :*   *.      :

NTHi_R2866     1.1   FDSFENINNSKEISFYDKE-FTEDYLVG------F--AQGFGVAKRNGETEE-TVRQYFK  146
NTHi_86028NP   1.2   SVPFENIDNSKAISFYDQK-YTEDYLVG------F--AQGFGVAKRNGDTEE-PVRKYFK  147
NTHi_375        2    YYNDKPIDPDKKISFFDSP-YTPGYTAA------F--VQGFGVKERNGNTEE-QAKQFID  155
aeg_HE40       3.1   ------DSPKDDCQLFGSNTYEKKWFL-----TDFSE-QKYGAEEENGIFGNEYAKHFVY  152
aeg_HE07       3.2   ------SNDSDDCFLIGAPIPVLPAGIGLVGEDDFSDGDVIQPEKENGTWDAPNAKHFIL  159
                            . ..   : .                *    :     :.**       .:::.

NTHi_R2866     1.1   ECLN---TGKHSNDPTC-----KSYSTDAYNIKSDIFALN----------TIVKNSHINS  188
NTHi_86028NP   1.2   ECLN---TGNYNNDNCK-----ANPLASSVSIKSDIFTPR----------PTIKNSHINS  189
NTHi_375        2    EFRTRLKKGTYFHSLGFFYSDGMPYTWLYEFYKNDIFNSY----------KVVKHSNINS  205
aeg_HE40       3.1   PFRED----------IPFYSESEKLYKLPVLYQSYLYDNYETEKTSRPLNALINNRVYDF  202
aeg_HE07       3.2   PFQEK----------RNVTSNGKDV--PRTLFQSYLYSPI---HKKRPKNALIDGKVYDV  204
                                                          :. ::               :.    :

NTHi_R2866     1.1   EILAVGNYTKLMLSA-----QH-SSTWAEHLYSNAELSLTVQDNSHVIGQTIDLGALILT  242
NTHi_86028NP   1.2   EILAVGNYTELMLAA-----QPSASSWAEHLYSDPGLSLTVQDNSHVIGQTIDLGFLQLT  244
NTHi_375        2    EILAVGNVADLYYKHYIRSPIPLYADWSIL--TLPLPRLKVSENSHVIGQIIHLYRVDLE  263
aeg_HE40       3.1   NVLAIDNYRSKFPDE-----------------T-RGLTLTVKNQSEVRGATLQLLKMVLQ  244
aeg_HE07       3.2   DVLAIDNYRFKFPNE-----------------PLRTLTLTVQNRSEVRGATLQLWKMKLQ  247
                     ::**:.*                                  *.*.:.*.* *   :.*   :  *

NTHi_R2866     1.1   NSLWEPRWNSNIDYLATENADIRFNTKSESLLVKGNYAGGARF    285
NTHi_86028NP   1.2   NSLWEPRWNSNIDYLVTENAEIRFNTKNESLLVEGYYAGGARF    287
NTHi_375        2    NSLWEPRWDSDVSYLNLNNGHIRFNTKNDSLVVGESRIRPTPD    306
aeg_HE40       3.1   DSLWEPRFNSDVHHLETQNANIRFNSTNTRLTVHENYQGDGSR    287
aeg_HE07       3.2   DSLWEPRFNSDVHHLETQNANIRFNSTNTRLTVHENYQGDGSR    290
                     :******::*::  :*   :*..****:..   *  *
```