Two cases of type-a *Haemophilus influenzae* meningitis within the same week in the same hospital are phylogenetically unrelated but recently exchanged capsule genes

Yves Terrat[1], Lauge Farnaes[2], John Bradley[2], Nicolas Tromas[1] & B. Jesse Shapiro[1]*

1. Département de sciences biologiques, Université de Montréal, Canada
2. Rady Children's Hospital, University of California San Diego, USA

* correspondence: jesse.shapiro@umontreal.ca

ABSTRACT

H. influenzae causes common and sometimes severe pediatric disease including chronic obstructive respiratory disease, otitis media, and infections of the central nervous system. Serotype b strains, with a b-type capsule, have been the historical cause of invasive disease, and the introduction of a serotype b-specific vaccine has led to their decline. However, unencapsulated or non-b-type *H. influenzae* infections are not prevented by the vaccine and appear to be increasing in frequency. Here we report two pediatric cases of severe central nervous system *H. influenzae* infection presenting to the same hospital in San Diego, California during the same week in January 2016. Due to good vaccine coverage in this part of the world, *H. influenzae* cases are normally rare and seeing two cases in the same week was unexpected. We thus suspected a recent transmission chain, and possible local outbreak. To test this hypothesis, we isolated and sequenced whole genomes from each patient and placed them in a phylogenetic tree spanning the known diversity of *H. influenzae.* Surprisingly, we found that the two isolates (H1 and H2) belonged to distantly related lineages, suggesting two independent transmission events and ruling out a local outbreak. Despite being distantly related, H1 and H2 belong to two different lineages that appear to engage in frequent horizontal gene transfer (HGT), suggesting overlapping ecological niches. Together, our comparative genomic analysis supports a scenario in which an f-type ancestor of H2 arrived in North America around 2011 and acquired an a-type capsule by recombination (HGT) with a recent ancestor of H1. Therefore, as in other bacterial pathogens, capsule switching by HGT may be an important evolutionary mechanism of vaccine evasion in *H. influenzae*.

OUTCOME

Two cases of severe central nervous system *H. influenzae* infection occurred during the same week in the same hospital in San Diego, California – a region where such infections are usually very rare due to vaccine coverage. We thus suspected a local outbreak of an *H. influenzae* clone not covered by the vaccine. Using whole genome sequencing and phylogenetic analysis of two isolates (H1 and H2, one from each patient), we found that they were distantly related, rapidly ruling out a local outbreak and suggesting independent transmission events. This result highlights the potential for rapid global spread of non-vaccine *H. influenzae* strains. In this case, both H1 and H2 both encoded a-type capsules, whereas the vaccine targets b-type capsules. We also present comparative genomic evidence that a recent f-type ancestor of H2 acquired an a-type capsule locus from a recent ancestor of H1, and that this horizontal gene transfer (HGT) event likely happened in the past decade in North America, but probably not in the San Diego hospital. These results highlight the potential importance in HGT of the capsule locus in allowing *H. influenzae* to escape vaccine coverage.

DATA SUMMARY

H. influenzae H1 and H2 genome sequences have been deposited in NCBI under BioProject PRJNA534512.

INTRODUCTION

Haemophilus influenzae is traditionally classified into encapsulated or unencapsulated strains, with encapsulated strains being subdivided into serotypes (or types) a-f, each with a distinct type of polysaccharide capsule. Type-b has long been associated with invasive disease (Pittman, 1931) and has thus been a major vaccine target. Since the introduction of vaccine against type-b *H. influenzae*, a dramatic decrease of severe cases has been observed (Peltola, 2000). However, this drop in severe type-b infections was followed by an increase of acute infections caused by non-b-type (*i.e.* a, c, d, e, and f capsule types) and non-typeable (unencapsulated) strains (Ladhani et al., 2012; Headrick et al., 2018; Tsang et Ulanova, 2017; Giufrè et al 2017).

As a common surface antigen and vaccine target, the capsule is a target of diversifying selection and the capsule locus is a hotspot of recombination in bacterial pathogens including *Klebsiella pneumoniae* (Wyres et al., 2015), *Streptococcus pneumoniae* (Mostowy et al., 2017), and *Neisseria meningitidis* (Bartley et al., 2017) – but has been less thoroughly studied in *H. influenzae*. Many (but not all) natural *H. influenzae* isolates are competent for DNA uptake (Maughan et Redfield, 2009), and *H. influenzae* housekeeping and lipopolysaccharide genes are inferred to undergo relatively frequent recombination (Vos et Didelot, 2009). Capsule types tend to be associated with particular phylogenetic lineages of *H. influenzae*, leading to the assertion that capsule genes evolve clonally, with limited recombination (De Chiara et al., 2014). On the other hand, the capsule locus can be deleted naturally by recombination (Kroll et al., 1988, and the capsule locus is occassionally recombined among phylogenetically distant lineages (Musser et al., 1988). Thus, capsular recombination in *H. influenzae* appears to be relatively rare, but its impact on the evolution and epidemiology of *H. influenzae* infections could be substantial. For example, capsular switching could allow successful pathogenic lineages to evade vaccines and persist. Alternatively, if capsular recombination is limited, we would expect vaccine lineages to be replaced with other lineages, encoding different capsules.

Here we describe two *H. influenzae* genomes, each isolated from a meningitis patient at Rady Children's Hospital (California, USA) within one week of each other in January of 2016.

81 As such severe central nervous system infections are extremely rare in Western countries since
82 the introduction of *H. influenzae* vaccines in the 1990s, the appearance of two cases in a such
83 narrow geographic and time window lead us to address the following questions using a
84 comparative genomic approach:

85 1) Are the two strains closely related, suggesting an outbreak of a particular *H. influenzae*
86 clone – possibly a vaccine-escape mutant? By placing the two strains on a phylogeny
87 of other sequenced *H. influezae* genomes, we found that the two strains were
88 unrelated. This surprising result led us to ask a second question:
89 2) Do these two unrelated strains share particular genes that might have allowed them
90 both to emerge at the same place and time?

91

92 **METHODS**

93 **Strain collection and patient characteristics**

94 *H.influenzae* strains were isolated from blood culture of two unrelated individuals (Table 1).
95 Patient #1 was an 11-month old female while Patient #2 was a 4-year-old male. They presented to
96 Rady Children's Hospital within one week of each other in January 2016. They were both treated
97 with antibiotics and were eventually cured with no apparent complications. Blood culture was *H.*
98 *influenzae* positive for both patients and showed that these strains were non-type b, but with an
99 encapsulated appearance. Both strains were sent to the United States Centers for Disease Control
100 and Prevention (CDC) for serotyping, which confirmed them both to be type a. Further patient
101 characteristics are given in Table 1. Strains isolated from patients 1 and 2 were respectively
102 named H1 and H2 in this study.

103

104 **Table 1**. Patient characteristics.

	Patient #1	Patient #2
Age	11 months	4 years
Positive culture	Blood culture	Blood and Cerebrospinal fluid (CSF) culture
CSF cell profile on first tap	Protein 267 Glucose 47 Erythrocytes 112,	Protein 68 Glucose 49 Erythrocytes 28

	Nucleated cells 619	Nucleated cells 707
CNS complications	Subdural empyema	Seizure
Antibiotic Treatment	Vancomycin and meropenem for 18 days	Ceftriaxone for 10 days
Time to fever resolution	14 days	1 day

DNA extraction & sequencing

Haemophilus influenzae strains were cultured overnight on chocolate agar plates (Thermo Fischer Sceintific) and DNA was extracted using the Bactozol DNA extraction kit (MRC inc.). Extracted DNA was further purified using the PowerClean® Pro DNA Clean-Up Kit (MOBIO Laboratories Inc.). Libraries were prepared using the Nextera XT kit (Illumina Inc.) following the standard Illumina protocol and library size was confirmed at approximately 1000 bp with a Qiaxcel Advanced System (QIAGEN). We performed paired end sequencing (2 x 300 bp) using the MiSeq reagent Kit V3 (Illumina Inc.) on the MiSeq system (Illumina Inc.) yielding a total of 1,128,523 paired-end reads for H1 and 1,708, 296 paired-end reads for H2.

Genome assembly, annotation, and phylogenetics

Sequencing reads were trimmed with Trimmomatic (Bolger et al., 2014) with default parameters. Trimmed reads were assembled into contigs using IDBA (Peng et al., 2012). We then compared the H1 and H2 genomes with a dataset of 80 non-typable, six b-capsule and one f-capsule genomes (previously described by De Chiara et al. 2014) using gene-by-gene alignments, described below. Consistent with the previous analysis of De Chiara et al., we identified six well-supported clades (named I-VI, following their nomenclature). To ensure that we did not miss close relatives of H1 or H2 not present in the De Chiara dataset, we searched NCBI for closely related genomes using a set of universal single copy genes extracted from H1 and H2 (Creevey et al., 2011). Using this approach, we identified two additional recently sequenced closely related encapsulated genomes: one f-capsule type isolated in Sweden in 2011 (Resman et al., 2011 ; Su et al., 2013) and one a-capsule type isolated in Canada in 2011 (NCBI accession number: CP017811.1). Ten sequences of *Haemophilus haemolyticus* were used as an outgroup for phylogenetic analyses. Contigs were annotated using the RAST server (Aziz et al., 2008). Translated gene predicitions were assigned to orthologous groups using Orthofinder (Emms and

131 Kelly, 2015). 941 genes assigned to the core genome (present in single copy in each genome)
132 were aligned using MUSCLE (Edgar, 2004). Concatenation of these aligned genes was to infer a
133 core genome phylogeny using FastTreeMP (Price et al., 2009). We also used FastTreeMP to
134 reconstruct each individual gene tree (or gene fragment for genes in the capsule operon).
135 Phylogenies were displayed in FigTree (http://tree.bio.ed.ac.uk/software/figtree/). Gene
136 presence/absence heatmaps were displayed in R with the heatmap.2 function of the ggplots
137 package (http://www.R-project.org/). The same exact protocole of alignment and phylogenetic
138 inference was applied on the smaller regions of the capsule genes.
139
140 **Horizontal gene transfer detection**
141 To infer putative recent recombination due to horizontal gene transfer (HGT) between two
142 strains, we used an explicit phylogenetic method where individual gene trees (including both core
143 and flexible genes) were screened for phylogenetic incongruence with the reference tree,
144 containing six major lineages, based on 941 core genes. A phylogenetic incongruence was
145 considered as an HGT if a gene sequence from a lineage [i] was grouped in a well-supported
146 clade (bootstrap value of 100) with a gene from a different lineage [j]. We also required the gene
147 sequences from lineages [i] and [j] to share >97% nucleotide identity. As some flexible genes are
148 shared uniquely between two isolates from different clades, we also considered these as putative
149 HGTs if they shared >97% nucleotide identity.
150
151 **RESULTS**
152
153 **H1 and H2 genomes are distantly related**
154 The core genome phylogeny based on 941 aligned genes shows that H1 and H2 belong to two
155 distinct lineages: H1 in clade VI and H2 in clade I (**Figure 1**). These two clades are distantly
156 related, clade I being the first *H. influenzae* branch after its divergence from the *H. haemolyticus*
157 outgroup. Within clade VI, H1 is the only type-a isolate in a sub-clade carrying b-type capsule
158 genes, with the exception of one genome, Hi1008, which likely lost most of its capsule genes
159 (Meyler et al., 2019). We searched the NCBI genome database to identify close relatives of H1
160 and H2 not present in our phylogeny. The closest sequenced relative of H1, NML_Hia1, was
161 isolated in Canada in 2011, and also carries an a-type capsule locus. The closest sequenced

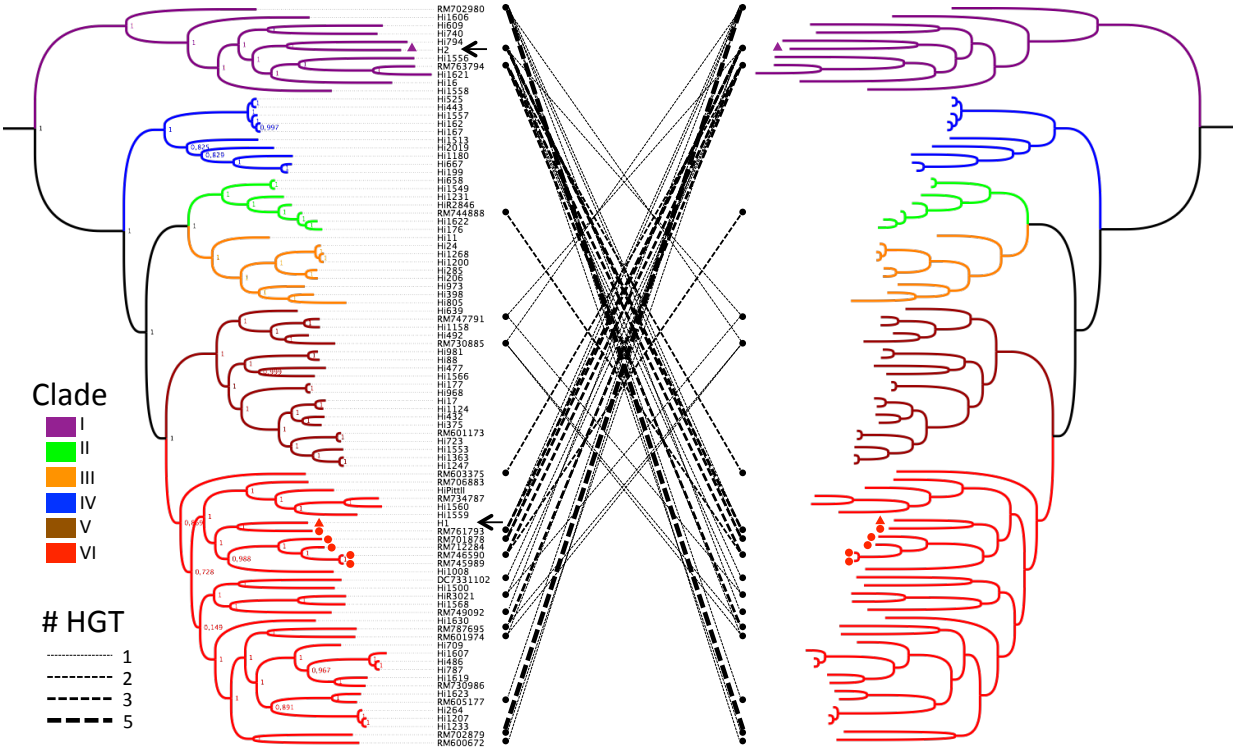
162 relatives of H2 are two f-type genomes: Hi794 (Finland, date unknown) and KR494 (Sweden,
163 2011). As H1 and H2 are distantly related in the tree, they are clearly not epidemiologically
164 linked and likely represent independent infection events.
165
166 **H1 and H2 belong to clades engaging in pervasive HGT**
167 Despite being phylogenetically unrelated, H1 and H2 were both isolated during the same week
168 from the same hospital. We thus hypothesized that they may have recently exchanged genes via
169 HGT. Using phylogenetic criteria to detect HGT (Methods) we did not identify any recent gene
170 transfers between H1 and H2 (**Figure 1**). Rather, H1 and H2 had flexible gene profiles similar to
171 other genomes from their respective clades (**Figure S1**). Despite the lack of recent HGT between
172 H1 and H2, we did observe that HGTs were unevenly distributed between clades (chi-square =
173 163.46, df = 14, $P < 2.2e-16$) and that the two distantly related clades containing H1 and H2
174 (clades I and VI) engaged in particularly pervasive HGT (**Figure 1**). These HGTs encode a mix
175 of virus- and plasmid-related functions, antibiotic resistance genes, and metabolic genes (**Table**
176 **S1**). Clades I and VI have similar profiles of selected virulence factors (De Chiara et al. 2014),
177 and members of both clades tend to encode *hap*, *hia*, and *hif* virulence genes (**Figure S2**). Neither
178 clade has a strong preference for a particular geographic region (**Figure S3**), time period (**Figure**
179 **S4**), or infection site (**Figure S5**). It is thus unclear why clades I and VI apparently engage in
180 more frequent HGT than other clades.
181
182 **Detailed phylogeny and HGT of the capsule locus**
183 The presence of two a-type strains (H1 and H2) in two distantly-related clades lead us to
184 investigate in greater details the evolution of capsule locus genes. By manually inspecting
185 individual gene alignments at this locus, we found that H1 and H2 had identical or very similar
186 sequences spanning a ~5kb region encompassing most of the serotype-specific genes (**Figure 2**).
187 These genetic similarities between H1 and H2 were not detected in the gene-by-gene analysis
188 (**Figure 1**) because of conflicting phylogenetic signals within genes (**Figure 2**). NML_Hia1,
189 another a-type genome isolated in Canada in 2011, also shared a similar or identical sequence in
190 the serotype-specific region, suggesting that the putative recombination event in this region
191 occurred in 2011 or earlier, and that the sequence has subsequently accumulated relatively few
192 mutations. Upstream and downstream of the serotype-specific region, H2 was most similar to f-

193 type strains (**Figure 2**). This suggests that an f-type ancestor of H2 acquired ~5kb of serotype-
194 specific DNA from an a-type donor strain, resulting in a serotype switch. Thus, recent ancestors
195 of H1 and H2 engaged in HGT at the capsule locus. However, the H1 and H2 are non-identical,
196 notably in the *acsC* gene (6 substitutions over 2 400 bp) (**Figure 2**), making it unlikely that the
197 exchange occured in Rady Children's Hospital.
198
199 **DISCUSSION**
200 The appearance of two *H. influenzae* infections in the same hospital in the same week was highly
201 unexpected because such infections are exceedingly rare in areas of high vaccine coverage. This
202 raised concerns that the two cases were part of a local outbreak, involving *H. influenzae*
203 transmission in the San Diego area. By sequencing the two isolate genomes (H1 and H2) and
204 placing them on a phylogenetic tree encompassing the known diversity of *H. influenzae*, we
205 found that they belonged to distantly related clades, indicating that each infection was aquired
206 independently and the two were not linked in a recent transmission chain. The entire analysis,
207 from strain isolate to sequencing and phylogenetic analysis, could be performed in about a week,
208 allowing us to rapidly rule out local transmission. Rather, the observation that H1 and H2 are
209 distantly related highlights the potential for rapid, potentially global spread of different *H.*
210 *influenzae* lineages.
211 To our knowledge H2 is the first North American clade I genome to be sequenced,
212 suggesting a recent transmission from Europe or Asia (**Figure S3**). H1 is part of clade VI and
213 shares recent ancestry with another type-a Canadian isolate from 2011, suggesting that this type-a
214 lineage has been circulating in North America for several years. Despite being distantly related,
215 H1 and H2 share a very similar capsule locus, particularly in the serotype-specific region.
216 Flanking the serotype-specific region and elsewhere in the genome, H2 is more similar to f-type
217 strains. Together, these observations point to a scenario in which an f-type ancestor of H2 arrived
218 in North America around 2011 and acquired an a-type capsule by recombination with a recent
219 ancestor of H1. We also note that H1 and H2 belong to two clades (VI and I, respectively) that
220 appear to engage in preferential HGT, possibly to cryptic niche overlap. Future work will be
221 needed to confirm and understand the reasons for this preferential genetic exchange.
222 That both H1 and H2 were isolated at the same place and time appears to be a
223 coincidence, but does suggest that these a-type strains are filling a vacant niche left by b-type

224 strains targeted by current vaccines. Our results also indicate that vaccines need not select for
225 lineage replacement, but could allow multiple different lineages to adapt via acquisition of new
226 capsule loci by HGT. We show that such a scenario is plausible, but further analysis of larger
227 population genomic samples will be needed to assess the relative importance of lineage
228 replacement vs. capsule HGT in the evolutionary response of *H. influenzae* to vaccine pressure.
229

230 **Figures and legends**

231



232

233

234 **Figure 1. Core genome phylogeny of *H.influenzae* and putative HGT events between clades.**
235 *H. haemolitycus* was used as an outgroup to root the tree (not shown). Encapsulated strains are
236 indicated with a circle (b-type) or a triangle (a-type); the remaining strains are unencapsulated.
237 H1 and H2 are indicated with arrows. Clades I-VI (following the nomenclature of De Chiara et al.
238 2014) are indicated in different colours. Putative HGT events between clades are indicated with
239 dashed lines. Line thickness indicates the number of HGTs (ranging from 1 to 5 genes).

240

241

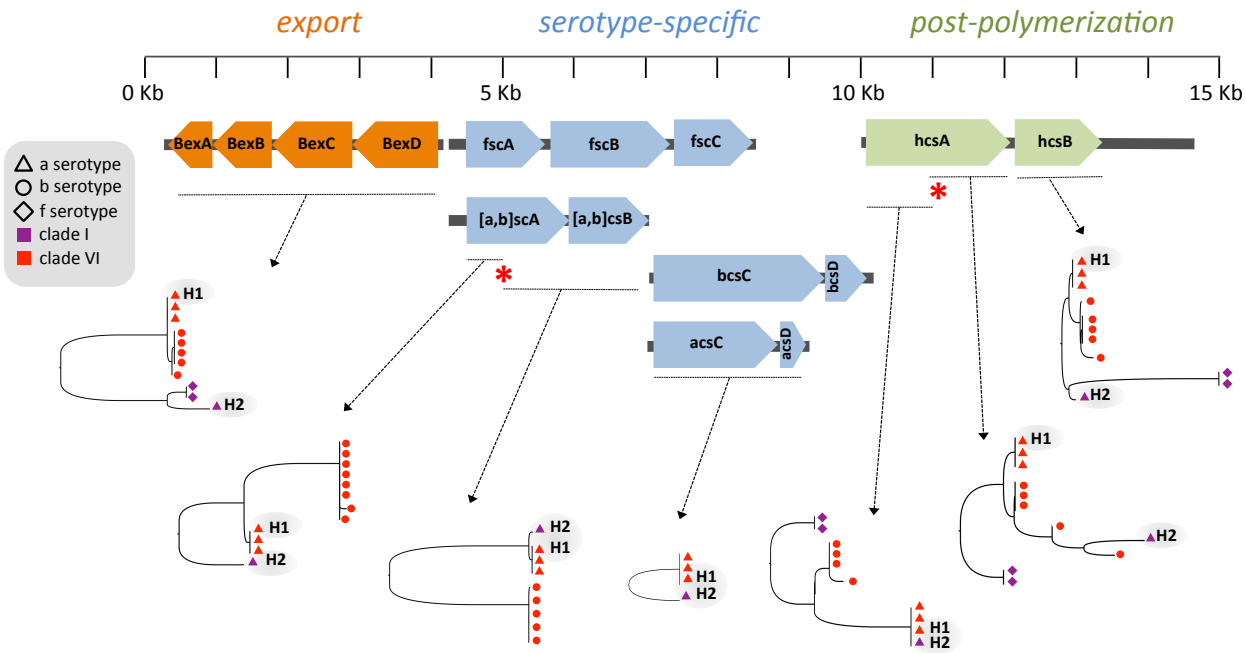


Figure 2: Phylogenies of genes and regions within the capsule locus. Dashed horizontal lines indicate the region used for phylogenetic analysis. Red stars indicate putative recombination breakpoints.

249 **Supplementary Table.**

250

251 **Table S1.** Annotated functions of putative HGTs between clades. Strain 1 and 2 indicated the two
252 strains involved in the putative HGT. Clades involved (I-VI) are shown in the far left column.

	Strain 1	Strain 2	Annotation
I vs VI	DC7331102	RM763794	D-hexose-6-phosphate mutarotase
	H2	RM601974	peptidase M16
	H2	RM701878	Bicyclomycin resistance protein
	H2	RM701878	EamA/RhaT family transporter
	H2	RM702879	Gamma-glutamylputrescine oxidoreductase
	H2	RM747791	6,7-dimethyl-8-ribityllumazine synthase
	NML_Hia1	H2	mannonate dehydratase
	RM600672	RM702980	iron-sulfur cluster insertion protein ErpA
	RM603375	RM702980	anticodon nuclease / ATPase AAA
	RM603375	RM702980	restriction endonuclease subunit S
	RM603375	RM702980	HD domain-containing protein - guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase
	RM603375	RM702980	NADH oxidase
	RM603375	RM702980	toxin Fic family protein
	RM603375	RM763794	electron transport complex subunit RsxE
	RM603375	RM763794	phage tail protein
	RM605177	RM702980	osmotically-inducible protein OsmY
	RM701878	RM702980	YfcC family protein
	RM702980	RM730885	nitrate/nitrite two-component system sensor histidine kinase NarQ
	RM702980	RM749092	TonB-dependent siderophore receptor
	RM702980	RM749092	phage baseplate protein
	RM702980	RM749092	phage tail protein
	RM702980	RM761793	long-chain-fatty-acid--CoA ligase FadD
	RM749092	RM763794	hypothetical protein
	RM761793	RM763794	Transcription termination/antitermination protein NusG
	RM763794	RM787695	plasmid RP4 TraN-related protein
	RM763794	RM787695	HD domain-containing protein / guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase
	RM763794	RM787695	hypothetical protein
III vs VI	RM744888	RM746590	hypothetical protein
	RM744888	RM746590	conjugative coupling factor TraD
V vs VI	HiR3021	RM747791	conjugal transfer protein TraG
	RM601974	RM730885	5'-methylthioadenosine/adenosylhomocysteine nucleosidase
	RM730885	RM787695	long-chain fatty acid--CoA ligase

253

254 **References :**

255

256 Aziz, R.K. *et al.* The RAST Server: rapid annotations using subsystems technology. *BMC*
257 *Genomics* **9**, 75 (2008).

258 Bartley, S.N. *et al.* Acquisition of the capsule locus by horizontal gene transfer
259 in Neisseria meningitidis is often accompanied by the loss of UDP-GalNAc synthesis. *Sci Rep* **7**,
260 44442 (2017).

261 Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina
262 sequence data. *Bioinformatics* **30**, 2114-20 (2014).

263 Creevey, C.J., Doerks, T., Fitzpatrick, D.A., Raes, J. & Bork P. Universally
264 distributed single-copy genes indicate a constant rate of horizontal transfer. *PLOS ONE* **6**, e22099
265 (2011).

266 De Chiara, M. *et al.* Genome sequencing of disease and carriage isolates of nontypeable
267 Haemophilus influenzae identifies discrete population structure. *Proc. Natl. Acad. Sci. U.S.A.*
268 **111**, 5439–5444 (2014).

269 Edgar, R. MUSCLE: multiple sequence alignment with high accuracy and high
270 throughput. *Nucleic Acids Res* **32**, 1792-1797 (2004).

271 Emms D.M & Kelly, S. OrthoFinder: solving fundamental biases in whole genome
272 comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16**, 157 (2015).

273 Giufrè, M., Cardines, R., Brigante, G., Orecchioni, F. & Cerquetti, M. Emergence of
274 Invasive Haemophilus influenzae Type A Disease in Italy. *Clin. Infect. Dis.* **64**, 1626–1628
275 (2017).

276 Headrick, A., Schmit, E. O. & Kimberlin, D. W. Fulminant Haemophilus Influenzae Type
277 a Infection in a 4-year-old with Previously Undiagnosed Asplenic Heterotaxy. *Pediatr. Infect.*
278 *Dis. J.* **37**, e108–e110 (2018).
279 Heikkinen, T. *et al.* Incidence of influenza in Finnish children. *Pediatr. Infect. Dis. J.* **22**,
280 S204-206 (2003).
281 Kroll, J. S., Hopkins, I. & Moxon, E. R. Capsule loss in H. influenzae type b occurs by
282 recombination-mediated disruption of a gene essential for polysaccharide export. *Cell* **53**, 347–
283 356 (1988).
284 Ladhani, S. N. *et al.* Invasive Haemophilus influenzae serotype e and f disease, England
285 and Wales. *Emerging Infect. Dis.* **18**, 725–732 (2012).
286 Maughan, H. & Redfield, R. J. Extensive Variation in Natural Competence in
287 Haemophilus Influenzae. *Evolution* **63**, 1852–1866 (2009).
288 Meyler, K. *et al.* Spontaneous capsule loss in Haemophilus influenzae serotype b
289 associated with Hib conjugate vaccine failure and invasive disease. *Clin Microbiol Infect* **25**, 390-
290 391 (2019).
291 Mostowy, R.J. *et al.* Pneumococcal capsule synthesis locus cps as evolutionary hotspot
292 with potential to generate novel serotypes by recombination. *Mol Biol Evol* **34**, 2537-2554
293 (2017).
294 Musser, J. M., Kroll, J. S., Moxon, E. R. & Selander, R. K. Evolutionary genetics of the
295 encapsulated strains of Haemophilus influenzae. *Proc Natl Acad Sci U S A* **85**, 7758–7762
296 (1988).
297 Peng, Y., Leung, H.C., Yiu, S.M. & Chin F.Y. IDBA-UD: a de novo assembler for single-
298 cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420-8
299 (2012).

300

301 Pittman, M. Variation and type specificity in the bacterial speces Hemophilus influenzae.
302 *J. Exp. Med.* **53**, 471–492 (1931).

303 Price, M.N., Dehal, P.S. & Arkin, A.P. FastTree: computing large minimum evolution
304 trees with profiles instead of a distance matrix. *Mol Biol Evol* **26**, 1641-50 (2009).

305 Resman, F. *et al.* Necrotizing myositis and septic shock caused by Haemophilus
306 influenzae type f in a previously healthy man diagnosed with an IgG3 and a mannose-binding
307 lectin deficiency. *Scand J Infect Dis* **43**, 972-6 (2011).

308 Su, Y.C., Hörhold, F., Singh, B. & Riesbeck, K. Complete Genome Sequence of
309 Encapsulated Haemophilus influenzae Type f KR494, an Invasive Isolate That Caused
310 Necrotizing Myositis. *Genome Announc* **1**, e00470-13 (2013).

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

314 Vos, M. & Didelot, X. A comparison of homologous recombination rates in bacteria and
315 archaea. *ISME J* **3**, 199–208 (2009).

316 Wyres, K.L *et al.* Extensive Capsule Locus Variation and Large-Scale Genomic
317 Recombination within the Klebsiella pneumoniae Clonal Group 258. *Genome Biol Evol* **7**, 1267-
318 79 (2015).