Comparison of the first whole genome sequence of '*Haemophilus quentini*' with two new strains of '*Haemophilus quentini*' and other *Haemophilus* species

Alasdair T. M. Hubbard[1] · Sian E. W. Davies[1] · Laura Baxter[2] · Sarah Thompson[3] · Mark M. Collery[1] · Daniel C. Hand[1] · D. John I. Thomas[1] · Colin G. Fink[1]

[1]*Micropathology Ltd., University of Warwick Science Park, Coventry, UK.*

[2]*School of Life Sciences, University of Warwick, Coventry, UK.*

[3]*Microbiology Department, Sheffield Teaching Hospitals NHS Foundation Trust, Northern General Hospital, Sheffield, United Kingdom*

Corresponding author: Dr. D. John I. Thomas, j.thomas@micropathology.com, 02476323222

Abstract

Comparison of the genome of the Gram negative human pathogen *Haemophilus quentini* MP1 with other *Haemophilus* species revealed that, although it is more closely related to *Haemophilus haemolyticus* than *Haemophilus influenzae*, the pathogen is in fact genetically distinct, a finding confirmed by phylogenetic analysis using the *H. influenzae* multilocus sequence typing genes. Further comparison with two other *H. quentini* strains recently identified in Canada revealed that these three genomes are more closely related than any other *Haemophilus* species, however there is still some sequence variation. There was no evidence of acquired antimicrobial resistance within the *H. quentini* MP1 genome nor any mutations within the DNA gyrase or topoisomerase IV genes known to confer resistance to fluoroquinolones, which has been previously identified in other *H. quentini* isolates. We hope by presenting the annotation and genetic comparison of the *H. quentini* MP1 genome it will aid the future molecular detection of this potentially emerging pathogen via the identification of unique genes that differentiate it from other *Haemophilus* species.

Keywords: *Haemophilus*; *Haemophilus quentini*; Whole genome sequencing; Annotation; Comparative genomics

Introduction

Haemophilus influenzae are clinically relevant Gram-negative bacteria which are known to cause a range of infections in humans, including sinusitis Brook et al. (2006) and conjunctivitis Van Dort et al. (2007), and most importantly neonatal sepsis Van Dcynse et al. (2016). Based on the varied ability to produce ornithine decarboxylase, urease and indole, *H. influenzae* can be separated into eight biotypes Andrzejczuk et al. (2017), as well as six serotypes based on their capsule, in addition to the strains that are non-encapsulated and therefore not typeable LaClaire et al. (2003). *H. influenzae* biotype IV are a group of non-typeable *H. influenzae* (NTHi) and are often characterised by the presence of peritrichous fimbriae Gousset et al. (1999); Quentin et al. (1989) which is thought to aid attachment in the urinary tract Gousset et al. (1999); Quentin et al. (1989). Although rare, *H. influenzae* biotype IV have been implicated as the cause of non-invasive infections in humans, including urogenital infections Alrawi et al. (2002); Harper and Tilse (1991); Quentin et al. (1989) and neonatal sepsis Van Dcynse et al. (2016); Wallace et al. (1983) and have been found to colonise the nasopharynx of children Jain et al. (2006) and the adult urinary tract Drouet et al. (1989). Importantly, ampicillin resistance within *H. influenzae* biotype IV has been observed Jain et al. (2006); Rashid et al. (2016). Initially DNA-DNA hybridisation and restriction fragment length polymorphism identified that at least a subset of *H. influenzae* biotype IV may actually be distinct, but still related, albeit distantly, from both *H. influenzae* and *Haemophilus haemolyticus* Quentin et al. (1993).Subsequent phylogenetic analysis of a partial sequence of the 16S rRNA gene of two strains of *H. influenzae* biotype IV found that a proportion of *H. influenzae* biotype IV were actually more closely related to *H. haemolyticus* than *H. influenzae* Quentin et al. (1996) and unofficially renamed '*Haemophilus quentini*', described as a "cryptic genospecies" since it cannot be distinguished phenotypically from *H. influenzae* Quentin et al. (1996). Since this discovery, it has been identified as the cause of urinary tract infections in men Glover et al. (2011) and neonatal bacteraemia through 16S rRNA sequencing Giufre et al. (2015); Hubbard et al. (2016); Mak et al. (2005). Importantly, *H. quentini* resistance to levofloxacin and tetracycline has been observed as well Mak et al. (2005). Aside from the description of these clinical cases involving *H. quentini*, there has been very little direct investigation into this possibly distinct pathogen.

Following the publication of the first whole genome sequence (WGS) of *H. quentini* Hubbard et al. (2016), we present the annotated assembly of the first WGS of *H. quentini* MP1. We have also compared the assembled genome to three other pathogenic *Haemophilus* species and two draft genomes of *H. quentini* that were

recently isolated in Canada Eshaghi et al. (2016) to determine if *H. quentini* is indeed closely related to *H. haemolyticus* and *H. influenzae* or whether it is in fact a distinct, novel, clinically relevant pathogen. Initially, *H. quentini* MP1 was identified via sequencing of the 16S rRNA extracted directly from clinical samples obtained from the cerebral spinal fluid (CSF) and joint fluid from a 9-day old male infant that presented with presumed sepsis in the UK Hubbard et al. (2016). The clinical isolate was subsequently obtained by culture from the CSF sample and it was reported to our laboratory that the isolate was susceptible to amoxicillin, co-amoxiclav, cefuroxime, ciprofloxacin and tetracycline by disc diffusion test performed during routine investigation in the NHS diagnostic laboratory. The clinical isolate was then subjected to whole genome sequencing in our laboratory Hubbard et al. (2016). We hope that by presenting the annotated WGS of *H. quentini* MP1, the future detection of this potentially emerging pathogen will be improved by the identification of genes that are unique to *H. quentini* which will allow it to be readily differentiated from other *Haemophilus* species in clinical molecular diagnostics.

Materials and methods

DNA extraction and sequencing

The *H. quentini* clinical isolate was provided to us on a nutrient agar slope by the Microbiology Department of Barnsley Hospital NHS Foundation Trust UK and the DNA extracted using an in-house method followed by WGS with an Illumina MiSeq. The bacterial isolate was re-suspended in 200 µl RNAse-free water, 10 mg/ml lysozyme and 40 µg/ml lysostaphin (all Sigma-Aldrich, UK), agitated at 37°C for 30 minutes at 500 rpm, then incubated at 72°C for 10 minutes with 200 µl ATL Buffer containing 40 µl protease (Qiagen, Germany). Extraction proceeded using High Pure Viral Nucleic Acid Kit (Roche, UK). Indexed library generation for WGS was performed using the Nextera DNA Library Preparation Kit and Nextera Index Kit (both Illumina, US). Pooled low-diversity 4 nM library with a 5% PhiX spike-in was sequenced with MiSeq 150 cycle Reagent Kit v3 (Illumina, US). Sequencing generated ~1.36 million 2 x 76 bp paired end reads.

Genome assembly

The quality of the raw MiSeq reads of the *H. quentini* MP1 genome were assessed using FastQC (version 0.11.5) and found to be high quality with no adapters present. Any sequencing reads that were found to be below a quality threshold of 30 were filtered out using FastQ Toolkit (version 2.2.0). Finally, *de novo* assembly of the *H. quentini* MP1 genome was performed using SPAdes(version 3.9.0) Bankevich et al. (2012) and the statistical analysis of the assembly were measured using QUAST(version 4.3) Gurevich et al. (2013)

Genome annotation

To produce a consensus sequence of the *H. quentini* MP1 genome, the contigs produced by *de novo* assembly were ordered and re-orientated against the reference genome *H. haemolyticus* M19107 Jordan et al. (2011) using *Mauve* (version 2.4.0) Darling et al. (2010). The final WGS of *H. quentini* MP1 was fully annotated using a combination of Prokaryotic Genome Annotation Pipeline (PGAP) Tatusova et al. (2014) and RAST (version 2.0) Aziz et al. (2008). RNAmmer Lagesen et al. (2007) was used to assess the presence of ribosomal RNA and Tandem Repeat Finder (version 4.09) Benson (1999) for the identification of the number of tandem repeats within the *H. quentini* MP1 genome.

Genomic comparisons

The WGS of *H. quentini* was compared to three other pathogenic *Haemophilus* species using a variety of comparative methods to establish the relative similarity or difference of the *H. quentini* genome and the relatedness of these four *Haemophilus* species. OrthoVenn Wang et al. (2015) was used to identify any overlapping or unique orthologous protein clusters in the genomes of *H. influenzae* PittGG Hogg et al. (2007), *H. haemolyticus* M19107 Jordan et al. (2011), *Haemophilus ducreyi* VAN5 Pillay et al. (2016), *H. quentini* MP1 Hubbard et al. (2016), *H. quentini* C860 and *H. quentini* K068 Eshaghi et al. (2016). The orthologous clusters were identified with default parameters, $1x10^{-5}$ E-value cut-off for all protein similarity comparisons and 1.5 inflation value for the generation of orthologous clusters.

To determine the relatedness of *H. quentini* MP1 and *H. influenzae* PittGG Hogg et al. (2007), *H. haemolyticus* M19107 Jordan et al. (2011) and *H. ducreyi* VAN5 Pillay et al. (2016) we calculated the average nucleotide identity (ANI) percentage identity using pyani (https://github.com/widdowquinn/pyani) between each of the genomes.

Finally, for a whole genome comparison of the *H. quentini* MP1 genome with the three other *Haemophilus* species, *H. quentini* MP1 was aligned with *H. influenzae* PittGG Hogg et al. (2007), *H. haemolyticus* M19107 Jordan et al. (2011), *H. quentini* C860 and *H. quentini* K068 Eshaghi et al. (2016) and alignment statistics were produced using progressiveMauve Darling et al. (2010).

Phylogenetic analysis

Phylogenetic analysis of the multilocus sequence typing (MLST) genes and 16S rRNA was performed to further ascertain how closely related the *H. quentini* MP1 is to three other Haemophilus species. The public MLST scheme for *H. influenzae* was selected for phylogenetic analysis of *Haemophilus* genomes, comprising genes

adk, atpG, frdB, mdh, pgi, and *recA* Meats et al. (2003). Gene sequences were extracted with NCBI BLAST+ toolkit in Galaxy Camacho et al. (2009); Cock et al. (2015) from the *H. quentini* MP1, *H. influenzae* PittGG Hogg et al. (2007), *H. influenzae* KW20 Gmuender (2001), *H. haemolyticus* M19107 Jordan et al. (2011), *H. ducreyi* VAN5 Pillay et al. (2016), *H. quentini* C860 and *H. quentini* K068 Eshaghi et al. (2016) genomes. Gene sequences were aligned separately and in concatemers using ClustalW Thompson et al. (1994). Evolutionary analyses were performed using MEGA7 Kumar et al. (2016). The Tamura-Nei (+G) model Tamura and Nei (1993) of maximum likelihood was selected using the in-built MEGA7 Bayesian Information Criterion (BIC) assessment of evolutionary models, and the bootstrap consensus phylogenetic tree was inferred from 500 replications Felsenstein (1985) (Fig. 3a). A further phylogenetic tree was constructed using the 16S rRNA sequences from these genomes, using the Hasegawa-Kishino-Yano model (+G +I) Hasegawa et al. (1985) of maximum likelihood in MEGA7 Kumar et al. (2016) using the methods described above (Fig. 3b).

Antimicrobial resistance genes

As *H. quentini* is potentially a novel pathogen, we tried to identify any possible antimicrobial resistance genes present in the genome. Antimicrobial resistance genes within the *H. quentini* MP1 gene were searched for using *RESfinder* (version 2.1) Zankari et al. (2012), a webtool that utilises BLAST Zhang et al. (2000) to identify antimicrobial resistance genes acquired by horizontal gene transfer using a database of known antimicrobial resistance genes. The search was performed with an identification threshold of 90% and minimum gene length of 80%. The protein sequence DNA gyrase subunits A and B and topoisomerase IV subunit A and B from *H. quentini* MP1 were aligned with the protein sequence of a multispecies *Haemophilus* equivalent to identify any mutations that may confer resistance to fluoroquinolones using ClustalOmega Sievers et al. (2011).

Results and discussion

Following *de novo* assembly of the *H. quentini* MP1 genome, any contigs that were found to be below <500 bp in length, <10X coverage or were identified to be non-bacterial when searched using BLAST Zhang et al. (2000) were manually removed. The finalised list of contigs produced a genome that was 2,151,950 bp (96x coverage) in length in 78 contigs with a N_{50} 67,683 bp and L_{50} 11 and a G+C content of 38.54%. The largest contig was 216,467 bp in length.

The annotated *H. quentini* MP1 genome, performed by PGAP Tatusova et al. (2014), identified a total of 2241 genes, of which 2187 were coding sequences (CDS), including 137 pseudo genes, and 54 RNA genes. Using Tandem Repeats Finder, 68 tandem repeats were found within the genome and RAST Aziz et al. (2008) did not

identify any transposable elements. The ribosomal RNA in the genome was also annotated, which found two 5S rRNA subunits, one 23S and one 16S rRNA subunit present within the genome. The identity of the assembled genome was confirmed as *H. quentini* through a BLAST search of the 16S rRNA subunit, which returned a positive match to an *H. quentini* 16S rRNA partial sequence, the same method by which recent infections with *H. quentini* Hubbard et al. (2016) have been identified.

Previous phylogenetic analysis of the sequenced 16S rRNA gene suggested that *H. quentini* was more closely related to *H. haemolyticus* than *H. influenzae* Quentin et al. (1996). However, this was not fully determined beyond the initial 16S rRNA comparison. To confirm whether this was indeed the case beyond the 16S rRNA gene, we compared the whole genome assembly of *H. quentini* MP1 to the genome of three other *Haemophilus* species. Both the *H. haemolyticus* M19107 Jordan et al. (2011) and *H. quentini* MP1 assemblies are fragmented, 123 and 78 contigs respectively, whilst *H. influenzae* PittGG Hogg et al. (2007) and *H. ducreyi* VAN5 Pillay et al. (2016) are each represented by a single chromosome. For this reason the contigs of the *H. haemolyticus* M19107 Jordan et al. (2011) and *H. quentini* MP1 assemblies were both re-ordered and oriented relative to *H. influenzae* PittGG Hogg et al. (2007) as the closest, complete, reference genome, using MAUVE Contig Mover Rissman et al. (2009). The ordered assemblies were then aligned to *H. influenzae* PittGG Hogg et al. (2007) and alignment statistics were calculated using progressiveMAUVE Darling et al. (2010). While it is clear from the alignment that there is sequence similarity between the genomes of *H. quentini* MP1 and *H. influenzae* PittGG Hogg et al. (2007) and *H. haemolyticus* M19107 Jordan et al. (2011), as shown by the segments of the aligned genome (Supporting information; Fig. S1a and Fig. S1b), there is also a considerable variation. This is visibly represented by a large number of gaps within the alignment (Supporting information; Fig. S1a and Fig. S1b) and is particularly evident by the presence of 168457 SNPs between the *H. quentini* MP1 and *H. influenzae* PittGG Hogg et al. (2007) genomes, while 151655 SNPs were present in the alignment of *H. quentini* MP1 and *H. haemolyticus* M19107 Jordan et al. (2011) genomes.

To further compare the four *Haemophilus* genomes beyond aligning the sequences, we performed a multi-species comparison of the shared orthologous protein clusters from genomes of *H. quentini* MP1, *H. influenzae* PittGG Hogg et al. (2007), *H. haemolyticus* M19107 Jordan et al. (2011) and *H. ducreyi* VAN5 Pillay et al. (2016). As was found with the sequence alignment, it was clear that *H. quentini* MP1 is more closely related to *H. haemolyticus* M19107 Jordan et al. (2011) than *H. influenzae* PittGG Hogg et al. (2007), with 145 and 101 overlapping orthologous protein clusters respectively. However, *H. quentini* MP1 contains 27 orthologous

204 protein clusters that are unique to the organism (Fig. 1a) and the number of proteins with no orthologous
205 match (278, *H. influenzae* PittGG Hogg et al. (2007); 154, *H. haemolyticus* M19107 Jordan et al. (2011); 490, *H.*
206 *quentini* MP1; 322, *H. ducreyi* VAN5 Pillay et al. (2016)) suggests there is a substantial difference among *H.*
207 *quentini* MP1 and other *Haemophilus* species. Finally, in order to examine the relatedness between *H. quentini*
208 MP1 and *H. influenzae* PittGG Hogg et al. (2007), *H. haemolyticus* M19107 Jordan et al. (2011) and *H. ducreyi*
209 VAN5 Pillay et al. (2016), the ANI between each of the genomes was calculated. Our results again confirm that
210 *H. ducreyi* VAN5 Pillay et al. (2016) is less similar to the other strains (0.84%, Fig. 2) and *H. quentini* MP1 is
211 more similar to *H. haemolyticus* M19107 Jordan et al. (2011), 0.95%, than *H. influenzae* PittGG Hogg et al.
212 (2007), 0.92%. Therefore, following whole genome analysis of *H. quentini* MP1 in comparison to other
213 *Haemophilus* species we found that it was indeed more similar to *H. haemolyticus* than *H. influenzae*.
214 However, despite previously being categorised as *H. influenzae* biotype IV there was still sufficient variation
215 between the genomes to suggest the *H. quentini* genome is distinct from the other *Haemophilus s*pecies
216 entirely.
217 Following the publication of the draft genome of *H. quentini* MP1 Hubbard et al. (2016), two other strains of *H.*
218 *quentini* were identified and sequenced in Canada and denoted strains C860 (accession number
219 MDJC00000000) and K068 (accession number MDJB00000000, Eshaghi et al. (2016)). Although *H. quentini* MP1
220 was more closely related to *H. haemolyticus* than any of the other *Haemophilus* species, following ANI
221 percentage analysis *H. quentini* MP1 was found to be much more closely related to *H. quentini* C860, 0.999%,
222 and *H. quentini* K068, 0.999%, than any other *Haemophilus* species (Fig. 2). However, alignment of *H. quentini*
223 MP1 with the two strains from Canada (Supporting information; Fig. S1c) highlighted that although all three
224 strains are very similar, the presence of 1828 and 1844 SNPs and 69 and 58 gaps when MP1 was aligned with
225 C860 and K068 respectively, suggests that there is still a large amount of variation between MP1 and the two
226 Canadian strains.
227 In contrast to the comparison of orthologous protein clusters of *H. quentini* MP1 with other *Haemophilus*
228 species, the three strains of *H. quentini* are very closely related and contain 2152 overlapping orthologous
229 protein clusters (Fig. 1b) confirming the ANI percentage analysis. However, again, the variation between the
230 three strains is evident with a number of proteins with no orthologous match (55; *H. quentini* MP1, 11; *H.*
231 *quentini* C860, 8; *H. quentini* K068). All three *H. quentini* strains also formed a distinct clade on both the MLST
232 and 16S phylogenetic tree, clearly separated from *H. influenzae* by high confidence bootstrap values, further

suggesting that *H. quentini* is in fact its own distinct species and not a biotype of *H. influenzae* (Fig. 3a and 3b). The two phylogenetic trees display an identical topology and concord with the genomic distance relationships suggested by the average nucleotide percentage identity comparison (Fig. 2). These analyses support the previous suggestion that *H. quentini* is more closely related to *H. haemolyticus* than *H. influenzae* by 16S rRNA sequencing, and that *H. ducreyi* VAN5 Pillay et al. (2016) was the least related to the three *H. quentini* genomes.

Ampicillin resistance in *H. Influenzae* biotype IV has previously been described Jain et al. (2006); Rashid et al. (2016), yet no acquired antibiotic resistance genes were identified in the *H. quentini* MP1 genome using RESfinder. We also found that the DNA gyrase subunit A and topoisomerase IV subunit A and B protein sequences shared 100% homogeneity with the corresponding multispecies *Haemophilus* protein sequences (accession number WP_005643730.1, WP_005642010.1 and WP_005642011.1, respectively). However, the DNA gyrase subunit B only shared 98.8% homogeneity with the protein sequence from DNA gyrase subunit B from multispecies *Haemophilus* (accession number WP_014550363.1), with two mutations at position 561 from Thr to Met (T561M) and 771 from Ile to Thr (I771T). The two mutations were also present in the two Canadian strains of *H. quentini*, though the mutations have not been identified as part of the quinolone resistance determining region Shoji et al. (2014) of *gyrB*. Therefore, this analysis confirmed the reported susceptibility to both ciprofloxacin and ampicillin found during routine laboratory investigations by the NHS Microbiology department. However, we were not able to confirm this due to being unable to perform phenotypic analysis on *H. quentini* MP1 isolate as a result of the limitations of our laboratory and access to insufficient sample.

In this paper, we have presented a novel genome that, although more closely related to *H. haemolyticus* than any of the other *Haemophilus* species compared, does not closely align with the three *Haemophilus* species. The distinct difference in genomes compared to *H. influenzae* and *H. haemolyticus* and the closer relatedness to two other *H. quentini* isolated in Canada raises questions as to whether *H. quentini* is actually a part of *H. influenzae* biotype IV or whether it is in fact a distinct, novel pathogen. Unique genes in the *H. quentini* MP1 genome can be exploited for molecular detection to differentiate *H. quentini* from other *Haemophilus* species, including *H. influenzae*, resulting in surveillance, improved diagnosis and treatment of infections caused by *H. quentini*.

Abbreviations: ANI; Average Nucleotide Identity, CDS; Coding DNA Sequence, HiB; *Haemophilus influenzae* Type B, MLST; Multilocus Sequence Typing, NTHI; Non-Typeable *Haemophilus influenzae*, PGAP; Prokaryotic Genome Annotation Pipeline, WGS; Whole Genome Sequencing

Accession number: This whole-genome shotgun project has been deposited at GenBank under the accession no. MCII00000000. The version described in this paper is version MCII02000000.

https://www.ncbi.nlm.nih.gov/nuccore/MCII00000000

Author Statements

Acknowledgements

We would like to thank the Microbiology Department of Barnsley Hospital NHS Foundation Trust UK for providing us with the bacterial isolate.

Conflicts of interest

We declare that there are no conflicts of interest.

References

Alrawi, A.M., Chern, K.C., Cevallos, V., Lietman, T., Whitcher, J.P., Margolis, T.P., et al. 2002. Biotypes and serotypes of Haemophilus influenzae ocular isolates Br J Opthtalmol. **86**(3): 276-277.

Andrzejczuk, S., Kosikowska, U., Malm, A., Chwiejczak, E., and Stepien-Pysniak, D. 2017. Phenotypic diversity of Haemophilus influenzae and Haemophilus parainfluenzae isolates depending on origin and health condition. Current Issues in Pharmacy and Medical Sciences, **30**(2): 90-99. doi: 10.1515/cipms-2017-0018.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., et al. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics, **9**(75). doi: 10.1186/1471-2164-9-75.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. **19**(5): 455-477. doi: 10.1089/cmb.2012.0021.

Benson, G. 1999. Tandem Repeats Finder: a program to analyze DNA sequences. Nucleic Acids Res. **27**(2): 573-580.

Brook, I., Foote, P.A., and Hausfeld, J.N. 2006. Frequency of recovery of pathogens causing acute maxillary sinusitis in adults before and after introduction of vaccination of children with the 7-valent pneumococcal vaccine. J Med Microbiol. **55**(Pt7): 943-946. doi: 10.1099/jmm.0.46346-0.

289 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. 2009. BLAST+: architecture
290 and applications. BMC Bioinformatics, **10**(421). doi: 10.1186/1471-2105-10-421.
291 Cock, P.J., Chilton, J.M., Gruning, B., Johnson, J.E., and Soranzo, N. 2015. NCBI BLAST+ integrated into Galaxy.
292 Gigascience, **4**(39). doi: 10.1186/s13742-015-0080-7.
293 Darling, A.E., Mau, B., and Perna, N.T. 2010. progressiveMauve: multiple genome alignment with gene gain,
294 loss and rearrangement. PLoS One, **5**(6): e11147. doi: 10.1371/journal.pone.0011147.
295 Drouet, E.B., Denoyel, G.A., Boude, M.M., Boussant, G., and de Montclos, H.P. 1989. Distribution of
296 Haemophilus influenzae and Haemophilus parainfluenzae Biotypes Isolated from the Human Genitourinary
297 Tract. European Journal of Clinical Microbiology and Infectious Diseases, **8**(11): 951-955.
298 Eshaghi, A., Soares, D., Tsang, R., Richardson, D., Kus, J.V., and Patel, S.N. 2016. Draft Genome Sequences of
299 Two "Haemophilus quentini" Isolates Recovered from Two Different Patients' Blood Cultures. Genome
300 Announc. **4**(6): e01321-01316. doi: 10.1128/genomeA.01321-16.
301 Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution, **39**(4): 783-
302 791. doi: 10.1111/j.1558-5646.1985.tb00420.x.
303 Giufre, M., Cardines, R., Degl'Innocenti, R., and Cerquetti, M. 2015. First report of neonatal bacteremia caused
304 by "Haemophilus quentini" diagnosed by 16S rRNA gene sequencing, Italy. Diagn Microbiol Infect Dis. **83**(2):
305 121-123. doi: 10.1016/j.diagmicrobio.2015.05.019.
306 Glover, W.A., Suarez, C.J., and Clarridge, J.E., 3rd. 2011. Genotypic and phenotypic characterization and clinical
307 significance of 'Haemophilus quentini' isolated from the urinary tract of adult men. J Med Microbiol. **60**(Pt11):
308 1689-1692. doi: 10.1099/jmm.0.031591-0.
309 Gmuender, H. 2001. Gene Expression Changes Triggered by Exposure of Haemophilus influenzae to Novobiocin
310 or Ciprofloxacin: Combined Transcription and Translation Analysis. Genome Research, **11**(1): 28-42. doi:
311 10.1101/gr.157701.
312 Gousset, N., Rosenau, A., Sizaret, P.Y., and Quentin, R. 1999. Nucleotide Sequences of Genes Coding for
313 Fimbrial Proteins in a Cryptic Genospecies of Haemophilus spp. Isolated from Neonatal and Genital Tract.
314 Infections Infection and Immunity, **67**(1): 8-15.
315 Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. 2013. QUAST: quality assessment tool for genome
316 assemblies. Bioinformatics, **29**(8): 1072-1075. doi: 10.1093/bioinformatics/btt086.

317 Harper, J.J., and Tilse, M.H. 1991. Biotypes of Haemophilus influenzae That Are Associated with Noninvasive
318 Infections. J Clin Microbiol. **29**(11): 2539-2542.
319 Hasegawa, M., Kishino, H., and Yano, T. 1985. Dating the human-ape split by a molecular clock of
320 mitochondrial DNA. Journal of Molecular Evolution, **22**: 160-174.
321 Hogg, J.S., Hu, F.Z., Janto, B., Boissy, R., Hayes, J., Keefe, R., et al. 2007. Characterization and modeling of
322 Haemophilus influenzae core and supragenomes based on the complete genomic sequences of Rd and 12
323 nontypeable strains. Genome Biology, **8**(6): R103. doi: 10.1186/gb-2007-8-6-r103).
324 Hubbard, A.T.M., Davies, S.E., Baxter, L., Thompson, S., Collery, M.M., Hand, D.C., et al. 2016. Draft Whole-
325 Genome Sequence of a Haemophilus quentini Strain Isolated from an Infant in the United Kingdom. Genome
326 Announc. **4**(5): e01075-01016. doi: 10.1128/genomeA.01075-16.
327 Jain, A., Kumar, P., and Awasthi, S. 2006. High ampicillin resistance in different biotypes and serotypes of
328 Haemophilus influenzae colonizing the nasopharynx of healthy school-going Indian children. J Med Microbiol.
329 **55**(Pt2): 133-137. doi: 10.1099/jmm.0.46249-0.
330 Jordan, I.K., Conley, A.B., Antonov, I.V., Arthur, R.A., Cook, E.D., Cooper, G.P., et al. 2011. Genome sequences
331 for five strains of the emerging pathogen Haemophilus haemolyticus. J Bacteriol. **193**(20): 5879-5880. doi:
332 10.1128/JB.05863-11.
333 Kumar, S., Stecher, G., and Tamura, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for
334 Bigger Datasets. Mol Biol Evol. **33**(7): 1870-1874. doi: 10.1093/molbev/msw054.
335 LaClaire, L.L., Tondella, M.L.C., Beall, D.S., Noble, C.A., Raghunathan, P.L., Rosenstein, N.E., et al. 2003.
336 Identification of Haemophilus influenzae Serotypes by Standard Slide Agglutination Serotyping and PCR-Based
337 Capsule Typing. J Clin Microbiol. **41**(1): 393-396. doi: 10.1128/jcm.41.1.393-396.2003.
338 Lagesen, K., Hallin, P., Rodland, E.A., Staerfeldt, H.H., Rognes, T., and Ussery, D.W. 2007. RNAmmer: consistent
339 and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35**(9): 3100-3108. doi: 10.1093/nar/gkm160.
340 Mak, G.C., Ho, P.L., Tse, C.W., Lau, S.K., and Wong, S.S. 2005. Reduced levofloxacin susceptibility and
341 tetracycline resistance in a clinical isolate of Haemophilus quentini identified by 16S rRNA sequencing. J Clin
342 Microbiol. **43**(10): 5391-5392. doi: 10.1128/JCM.43.10.5391-5392.2005.
343 Meats, E., Feil, E.J., Stringer, S., Cody, A.J., Goldstein, R., Kroll, J.S., et al. 2003. Characterization of Encapsulated
344 and Noncapsulated Haemophilus influenzae and Determination of Phylogenetic Relationships by Multilocus
345 Sequence Typing. J Clin Microbiol. **41**(4): 1623-1636. doi: 10.1128/jcm.41.4.1623-1636.2003.

Pillay, A., Katz, S.S., Abrams, A.J., Ballard, R.C., Simpson, S.V., Taleo, F., et al. 2016. Complete Genome Sequences of 11 Haemophilus ducreyi Isolates from Children with Cutaneous Lesions in Vanuatu and Ghana. Genome Announc. **4**(4): e00459-00416. doi: 10.1128/genomeA.00459-16.

Quentin, R., Martin, C., Musser, J.M., Pasquier-Picard, N., and Goudeau, A. 1993. Genetic Characterization of a Cryptic Genospecies of Haemophilus Causing Urogenital and Neonatal Infections. J Clin Microbiol. **31**(5): 1111-1116.

Quentin, R., Musser, J.M., Mellouet, M., Sizaret, P.Y., Selander, R.K., and Goudeau, A. 1989. Typing of Urogenital, Maternal, and Neonatal Isolates of Haemophilus influenzae and Haemophilus parainfluenzae in Correlation with Clinical Source of Isolation and Evidence for a Genital Specificity of H. influenzae Biotype IV. J Clin Microbiol. **27**(10): 2286-2294.

Quentin, R., Ruimy, R., Rosenau, A., Musser, J.M., and Christen, R. 1996. Genetic Identification of Cryptic Genospecies of Haemophilus Causing Urogenital and Neonatal Infections by PCR Using Specific Primers Targeting Genes Coding for 16S rRNA. J Clin Microbiol. **34**(6): 1380-1385.

Rashid, H., Shoma, S., and Rahman, M. 2016. Prevalence of β-Lactamase Positive Ampicillin Resistant H. Influenzae from Children of Bangladesh. Journal of Infectious Diseases and Epidemiology, **2**(2): 1-5.

Rissman, A.I., Mau, B., Biehl, B.S., Darling, A.E., Glasner, J.D., and Perna, N.T. 2009. Reordering contigs of draft genomes using the Mauve aligner. Bioinformatics, **25**(16): 2071-2073. doi: 10.1093/bioinformatics/btp356.

Shoji, H., Shirakura, T., Fukuchi, K., Takuma, T., Hanaki, H., Tanaka, K., et al. 2014. A molecular analysis of quinolone-resistant Haemophilus influenzae: validation of the mutations in Quinolone Resistance-Determining Regions. J Infect Chemother. **20**(4): 250-255. doi: 10.1016/j.jiac.2013.12.007.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. **7**(539). doi: 10.1038/msb.2011.75.

Tamura, K., and Nei, M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution, **10**(3): 512-526.

Tatusova, T., Ciufo, S., Fedorov, B., O'Neill, K., and Tolstoy, I. 2014. RefSeq microbial genomes database: new representation and annotation strategy. Nucleic Acids Res. **42**: D553-559. doi: 10.1093/nar/gkt1274.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**(22): 4673-4680.

Van Dcynse, E., Vermijlen, P., Van Noyen, G., and Van Noyen, R. 2016. Neonatal Sepsis Due to Nonencapsulated Haemophilus Influenzae Biotype IV. Acta Clinica Belgica, **52**(4): 204-206. doi: 10.1080/17843286.1997.11718577.

Van Dort, M., Walden, C., Walker, E.S., Reynolds, S.A., Levy, F., and Sarubbi, F.A. 2007. An outbreak of infections caused by non-typeable Haemophilus influenzae in an extended care facility. J Hosp Infect. **66**(1): 59-64. doi: 10.1016/j.jhin.2007.02.001.

Wallace, J., R. J.,, Baker, C.J., Quinones, F.J., Hollis, D.G., Weaver, R.E., and Wiss, K. 1983. Nontypable Haemophilus influenzae (Biotype 4) as a Neonatal, Maternal, and Genital Pathogen. Reviews of Infectious Diseases, **5**(1): 123-136.

Wang, Y., Coleman-Derr, D., Chen, G., and Gu, Y.Q. 2015. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res. **43**(W1): W78-84. doi: 10.1093/nar/gkv487.

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., et al. 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. **67**(11): 2640-26444. doi: 10.1093/jac/dks261.

Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. 2000. A Greedy Algorithm for Aligning DNA Sequences. Journal of Computational Biology, **7**(1-2): 203-214.

Figures and Tables

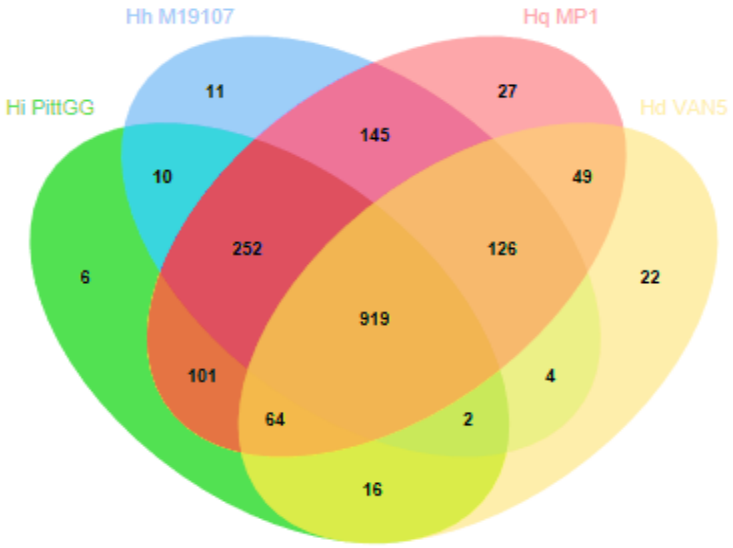
Figure 1: Venn diagram plotted with OrthoVenn shows the orthologous protein clusters among the genomes of (a) *Haemophilus quentini* (Hq MP1), *Haemophilus influenzae* PittGG (Hi PIttGG), *Haemophilus ducreyi* VAN5 (Hd VAN5) and *Haemophilus haemolyticus* M19107 (Hh M19107) and (b) *Haemophilus quentini* MP1 (Hq MP1), *Haemophilus quentini* C860 (Hq C860) and *Haemophilus quentini* K068 (Hq K068).

Figure 2: Average nucleotide identity percentage identity comparison of *H. quentini* MP1, *H. quentini* K068, *H. quentini* C860, *H. influenzae* PittGG, *H. ducreyi* VAN5 and *H. haemolyticus* M19107.

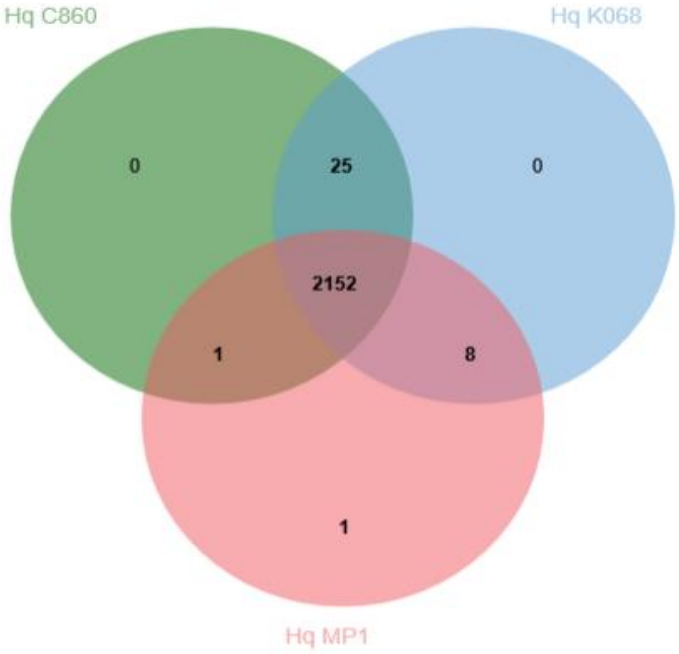
Figure 3: Phylogenetic analysis of (a) concatenated *adk, atpG, frdB, mdh, pgi,* and *recA* genes and the (b) 16S rRNA genes of *H. quentini* MP1 (Hq MP1), *H. quentini* K068 (Hq K068), *H. quentini* C860 (Hq C860), *H. influenzae* KW20 (Hinf KW20), *H. influenzae* PittGG (Hinf PittGG), *H. ducreyi* VAN5 (Hd VAN5) and *H. haemolyticus* M19107 (Hh M19107). The genes were aligned using ClustalW and the trees were produced using the (a) Tamura-Nei (+G) model of maximum likelihood and the (b) Hasegawa-Kishino-Yano model (+G +I) of maximum likelihood using MEGA7 with bootstrap calculations used to determine the confidence in the position of each clade on the tree.

Figures

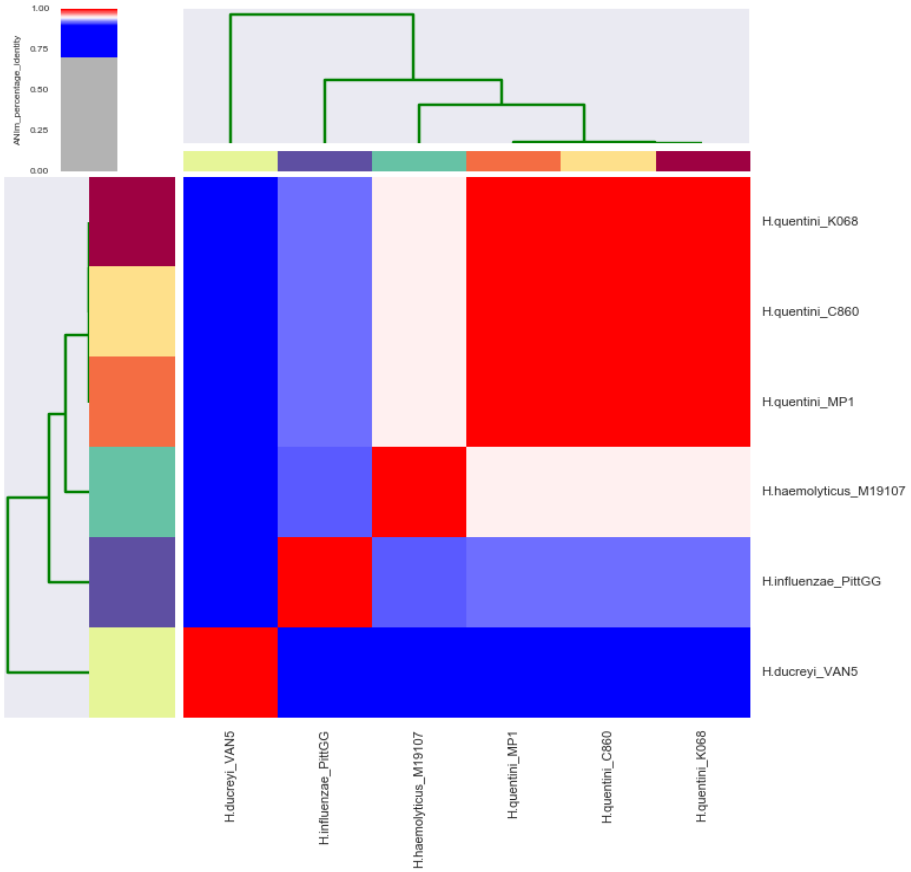
1(a)



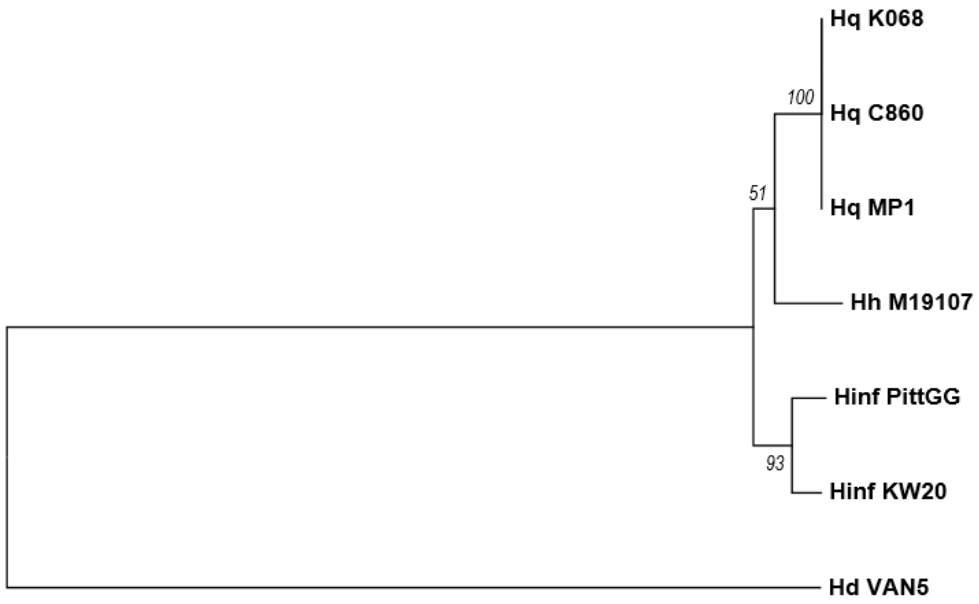
1(b)


Hq C860
Hq K068
Hq MP1
0
25
0
2152
1
8
1

2.

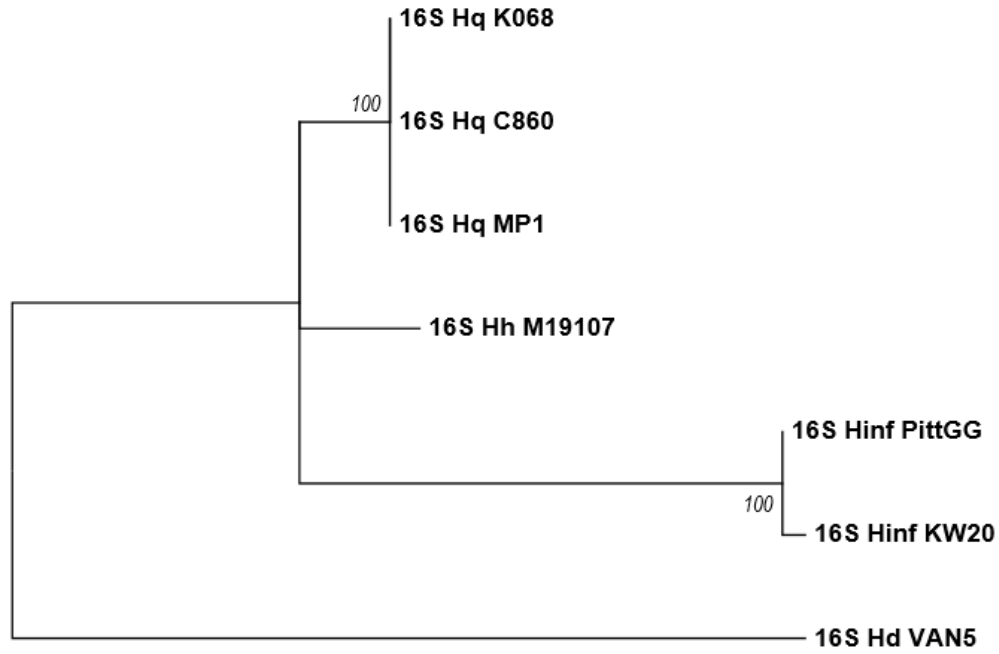

ANIm_percentage_identity
1.00
0.75
0.50
0.25
0.00
H.quentini_K068
H.quentini_C860
H.quentini_MP1
H.haemolyticus_M19107
H.influenzae_PittGG
H.ducreyi_VAN5
H.ducreyi_VAN5
H.influenzae_PittGG
H.haemolyticus_M19107
H.quentini_MP1
H.quentini_C860
H.quentini_K068

437 3(a).



438

439 3(b).



440

441

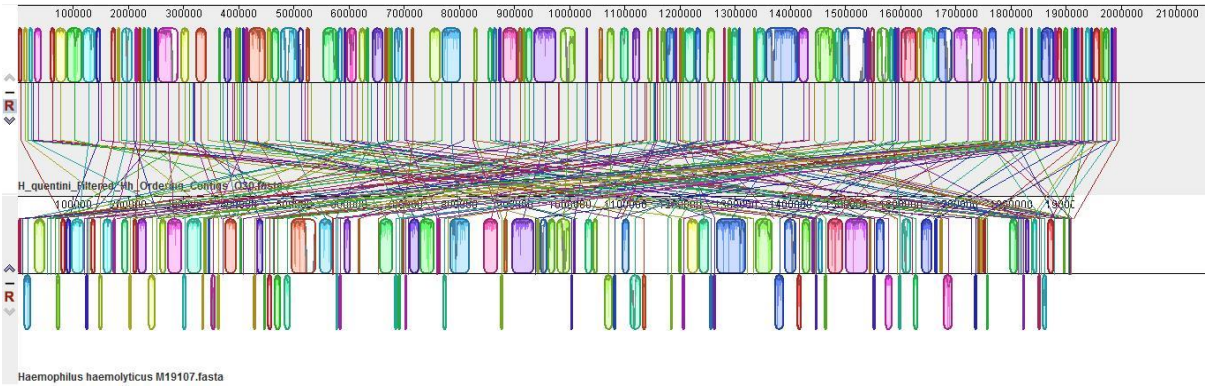
442

443

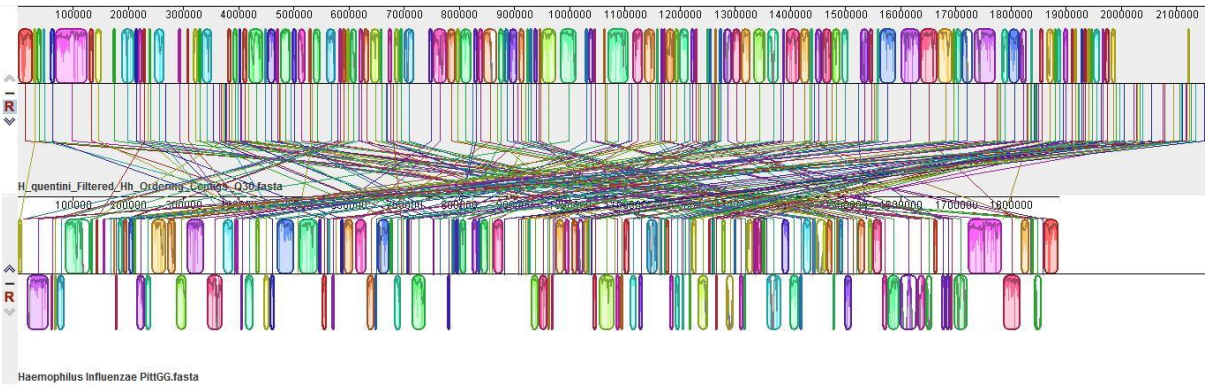
Supporting Information

gen-2017-0195.R1 Figure S1: Alignment of the *H. quentini* MP1 genome with the (a) *H. haemolyticus* M19107 and (b) *H. influenzae* PittGG reference genomes (c) *Haemophilus quentini* C860 and *Haemophilus quentini* K068. Locally Collinear Blocks are shown by boxes of the same colour and represent conserved sequence regions shared by one or more genomes. White areas represent unique sequence regions in the three genomes.

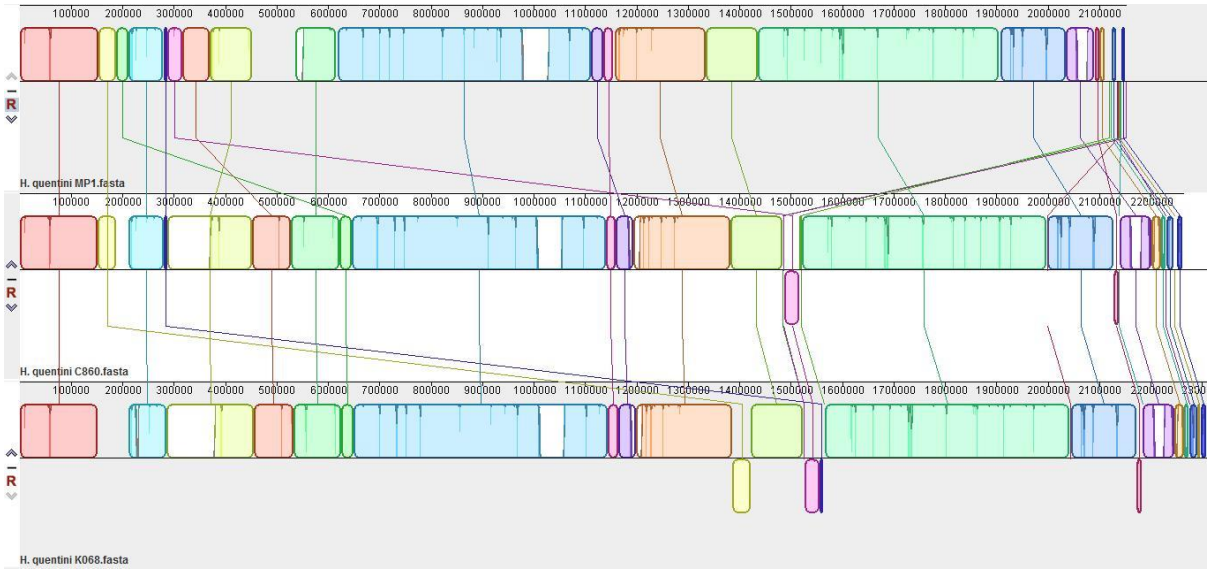
S1(a).



S1(b).



S1(c).


H. quentini MP1.fasta
H. quentini C860.fasta
H. quentini K068.fasta