Delineation of the Species *Haemophilus influenzae* by Phenotype, Multilocus Sequence Phylogeny, and Detection of Marker Genes^{∇†}

Niels Nørskov-Lauritsen,^{1*} Merete D. Overballe,¹ and Mogens Kilian²

Department of Clinical Microbiology, Aarhus University Hospital Skejby, DK-8200 Aarhus N,¹ and Institute of Medical Microbiology and Immunology, Aarhus University, DK-8000 Aarhus C,² Denmark

Received 4 June 2008/Accepted 26 November 2008

To obtain more information on the much-debated definition of prokaryotic species, we investigated the borders of *Haemophilus influenzae* by comparative analysis of *H. influenzae* reference strains with closely related bacteria including strains assigned to *Haemophilus haemolyticus*, cryptic genospecies biotype IV, and the never formally validated species “*Haemophilus intermedius*”. Multilocus sequence phylogeny based on six housekeeping genes separated a cluster encompassing the type and the reference strains of *H. influenzae* from 31 more distantly related strains. Comparison of 16S rRNA gene sequences supported this delineation but was obscured by a conspicuously high number of polymorphic sites in many of the strains that did not belong to the core group of *H. influenzae* strains. The division was corroborated by the differential presence of genes encoding *H. influenzae* adhesion and penetration protein, fuculokinase, and Cu,Zn-superoxide dismutase, whereas immunoglobulin A1 protease activity or the presence of the *iga* gene was of limited discriminatory value. The existence of porphyrin-synthesizing strains (“*H. intermedius*”) closely related to *H. influenzae* was confirmed. Several chromosomally encoded hemin biosynthesis genes were identified, and sequence analysis showed these genes to represent an ancestral genotype rather than recent transfers from, e.g., *Haemophilus parainfluenzae*. Strains previously assigned to *H. haemolyticus* formed several separate lineages within a distinct but deeply branching cluster, intermingled with strains of “*H. intermedius*” and cryptic genospecies biotype IV. Although *H. influenzae* is phenotypically more homogenous than some other *Haemophilus* species, the genetic diversity and multicluster structure of strains traditionally associated with *H. influenzae* make it difficult to define the natural borders of that species.

Bacterial diversity is usually expected to be organized into phenotypic and genetic clusters recognized as species. The diversity within species is thought to be confined by natural forces, although uncertainty exists as to the nature of the cohesion that constrains this diversity. There is increasing evidence that recombination occurs between distantly related bacteria and that barriers to this process, which could be used to define species naturally, are not always apparent (7, 11). As the bacterial species concept is central to taxonomy and evolutionary biology and has consequences for other disciplines such as the discrimination of pathogens from related bacteria with no pathogenic potential, there is a need to reevaluate to what extent accepted species are indeed coherent units.

The gram-negative, nonmotile, facultatively anaerobic bacterium *Haemophilus influenzae* is a commensal organism of the pharynx frequently involved in localized infections of the respiratory tract, middle ear, and conjunctiva and sometimes in invasive infections such as meningitis or bacteremia. Presumptive identification is based on inability to synthesize NAD (V factor dependence) and porphyrin/hemin (X factor dependence). The latter characteristic is particularly notable as it is

shared with few other microorganisms. Further metabolic characteristics of *H. influenzae* include fermentation of D-xylose, D-ribose, and D-galactose but not sucrose or D-mannose (17). By DNA analysis the X-factor-dependent species *H. influenzae*, *Haemophilus aegyptius*, and *Haemophilus haemolyticus* constitute a closely related group (4, 5, 14, 35), and housekeeping gene nucleotide dissimilarities between these species are exceeded by intraspecific nucleotide dissimilarities within the commensal species *Haemophilus parainfluenzae* (33).

H. aegyptius was conceived as a species distinct from *H. influenzae* with a particular propensity to cause conjunctivitis (37). However, the existence of *H. aegyptius* as a separate species is controversial (5), and the designation *H. influenzae* biogroup *aegyptius* has been proposed to include traditional *H. aegyptius* and the closely related clones responsible for the invasive infection known as Brazilian purpuric fever (3), although formally the epithet *aegyptius* has priority. *H. haemolyticus* is distinguished from *H. influenzae* primarily by its hemolytic action on erythrocytes. A recent characterization of a large number of presumptive *H. influenzae* isolates from the respiratory tract of patients with chronic obstructive airway disease demonstrated that 102 of 258 strains clustered with the type strain of *H. haemolyticus* by 16S rRNA gene sequence, although many of the isolates were nonhemolytic. In contrast to *H. influenzae*, the atypical isolates were not associated with exacerbations of the disease (31), emphasizing the need to discriminate between the two species.

Genetic methods have identified unnamed taxa of *Hae-*

* Corresponding author. Mailing address: Department of Clinical Microbiology, Aarhus University Hospital Skejby, Brendstrupgaardsvej, DK-8200 Aarhus N, Denmark. Phone: 45 8949 5603. Fax: 45 8949 5611. E-mail: nielnoer@rm.dk.

† Supplemental material for this article may be found at <http://jb.asm.org/>.

[∇] Published ahead of print on 5 December 2008.

mophilus related to *H. influenzae*. Variant strains isolated from the genitourinary tract were first reported from Canada (1) and have been further studied by Quentin and coworkers (13, 39). These “cryptic genospecies biotype IV” strains (negative for tryptophanase/indole production and positive for urease and ornithine decarboxylase) are distinguishable from typical strains of *H. influenzae* by multilocus enzyme electrophoresis, 16S rRNA gene sequence, and DNA hybridization. However, only some biotype IV strains belong to the cryptic genospecies, and the lack of phenotypic characters enabling discrimination from *H. influenzae* has been an obstacle to the formal validation as a species. Another taxon, “*Haemophilus intermedius*,” was suggested (but never validly published) by Burbach (4). By DNA hybridization and by selected phenotypic traits, these strains take up an intermediate position between *H. influenzae* and *H. parainfluenzae*. Two subspecies were described, “*Haemophilus intermedius* subsp. *intermedius*,” capable of synthesizing porphyrin from δ -aminolevulinic acid and of fermenting sucrose, and “*Haemophilus intermedius* subsp. *gazogenes*,” characterized by the fermentation of mannose and the production of gas from glucose.

A number of investigations have addressed the population structure of *H. influenzae* with assessment of particular virulence traits associated with specific subgroups or clades (9, 26, 32, 41). In the present study we attempted to define the borders of the species by performing phenotypic characterization, 16S rRNA sequence comparison, multilocus housekeeping gene sequence phylogeny, and detection of putative virulence and marker genes on a collection of 42 strains of *H. influenzae* and representatives of the above-mentioned closely related bacteria.

MATERIALS AND METHODS

Bacterial strains and phenotypic characterization. The 42 bacterial strains included in this study are listed together with information on origins and alternative strain designations in Table S1 in the supplemental material. The type strains of *H. influenzae* (HK389^T = NCTC 8143^T), *H. aegyptius* (HK 367^T = NCTC 8502^T), and *H. haemolyticus* (HK 386^T = NCTC 10659^T) were included together with seven reference strains of *H. influenzae* (HK 61, HK 1136 [=ATCC 9134, serotype b], HK 1210, HK 1212, HK 2067 [serotype f], HK 2122, and the full-genome-sequenced strain Rd, which is a rough variant of a serotype d encapsulated strain of biotype IV). Twenty strains classified as “*Haemophilus intermedius*” were obtained from the culture collection at Gothenburg University (CCUG) (www.ccug.se), four cryptic genospecies biotype IV strains were obtained from France (39) and Hong Kong (23), and five strains (three demonstrating hemolysis) isolated from sputum and assigned to *H. haemolyticus* on the basis of 16S rRNA gene sequence were obtained from the United States (31). Strain HK 855 was previously shown to carry an unusual *infB* sequence (14), PN24 is a recent isolate from urine demonstrating X factor independence, and the oral isolate HK 676 has been used as a representative of *H. haemolyticus* (14, 33).

Phenotypic profiles were obtained using the commercial test systems API NH and Vitek NH from BioMérieux. Generation of gas by fermentation of glucose, the oxidase (spot) test, and assessment of hydrogen sulfide emission using lead acetate paper were carried out as described previously (17). V-factor dependence was tested with tablets from Rosco Diagnostics placed on inoculated brain heart infusion agar plates. Hemolysis was evaluated on agar plates containing 5% horse blood. Synthesis of porphyrins from δ -aminolevulinic acid was assessed by a modification of the porphyrin test (16), as described previously (34). Immunoglobulin A1 (IgA1) protease activity was detected by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis of human myeloma IgA1 solution after overnight incubation with bacteria (27). Biotypes of *H. influenzae* were defined as described previously (18).

DNA sequencing. Fragments of the housekeeping genes *adh*, *atpG*, *frdB*, *fucK*, *mdh*, *pgi*, and *recA* were amplified by PCR and sequenced according to the

multilocus sequence typing (MLST) scheme (26). Primers *adh* 34f, *adh* 610r, *pgi* 1331r, and *recA* 54f (33) were used with selected strains giving weak or no amplification products with the MLST primers. The *atpG* gene in three strains, CCUG 11096, CCUG 24145, and CCUG 32367, encoded an isoleucine inserted between phenylalanine₁₃₁ and asparagine₁₃₂ (numbering with respect to strain Rd HI_0480). The corresponding triplet in *atpG* was excluded from the comparison. All other gene fragments were identical in length for all study strains. No *fucK* product could be amplified from 32 of the study strains, and this gene was omitted from the sequence comparison. The remaining six gene fragments were concatenated into a hybrid sequence of 2,712 nucleotides (nt) and used for analysis.

16S rRNA gene sequences were amplified by PCR and sequenced as described previously (20). A 1,362-nt fragment (1,361 nt in HK 2067 and HK 2122) corresponding to nt 27 to 1388 of the 16S rRNA gene in strain Rd was aligned and compared with the strain Rd sequence. The previously deposited sequences for the type strain of *H. aegyptius* (CCUG 25716^T, AY362905), the cryptic genospecies biotype IV strain S32F2 (AY962106), and strain Rd were used. *H. parainfluenzae* T3T1 (http://www.sanger.ac.uk/Projects/H_parainfluenzae/) was used as the outgroup.

Hemin biosynthesis genes *hemB* (prophobilinogen synthase, EC 4.2.1.24), *hemE* (uroporphyrinogen decarboxylase, EC 4.1.1.37), and *hemN* (coproporphyrinogen dehydrogenase, EC 1.3.99.22) from *Aggregatibacter actinomycetemcomitans* (<http://www.oralgen.lanl.gov/>), *Pasteurella multocida* (24), and *Histophilus somni* (6) were aligned, and a number of degenerate primers were designed. PCR amplification with primers *hemB* 284f (TATGGCRGAAGARGCYTA T)-*hemB* 818r (GTTTTTACACGATACACCATTA), *hemE* 58f (ATGACRC CSRTWTGGATGATGCG)-*hemE* 466r (CCATATAAGTAGCTAATGT), and *hemN* 219f (TACTCGTCATCAACATAAGG)-*hemN* 872r (GGTTTRG CRAARTGATCCAT) (numbering with respect to the corresponding genes in *A. actinomycetemcomitans*; i, deoxyinosine) resulted in products of the expected size, and sequencing revealed genes with a similarity in excess of 75% to corresponding genes from other representatives of the family Pasteurellaceae.

The translocase protein gene *tatC* is located 5' to *hemB* in *A. actinomycetemcomitans*, *P. multocida*, *H. somni*, and *H. parainfluenzae*, and a successful attempt to amplify the 5' end of *hemB* and flanking DNA was carried out using the upstream primer *tatC* 453f (TATTAGTAGTTATTTAGATTTT). The NADP-specific glutamate dehydrogenase gene (*gdhA*) is located 3' to *tatC* in *H. influenzae* (10) and 3' to *hemB* in *H. parainfluenzae* T3T1, and the 3' end of *hemB* and flanking DNA were amplified using the downstream primer *gdhA* 232r (TGTT TACTTGCACTTGCCCTTTAT) (numbering with respect to the corresponding genes in *H. influenzae* Rd).

The *hemE* gene is located between an NAD pyrophosphatase gene (*nudC*) and a hypothetical conserved gene (Clusters of Orthologous Groups of proteins 3068) in *A. actinomycetemcomitans*, *P. multocida*, *H. somni*, and *H. parainfluenzae*. The ends of *hemE* and flanking DNA were amplified using primers *nudC* 17f (AAGATGATTTTGGTTATTGG) and COG3068 286r (CATCTAACGCTGG CACAACACC) (*H. influenzae* Rd numbering).

Sequence analysis. The complete *H. influenzae* database was downloaded from the *Haemophilus* MLST website (<http://haemophilus.mlst.net/>) on 7 May 2007. After removal of the *fucK* fragment and addition of sequences from the present study strains, 45 sequences were redundant (42 sequence types [STs] were identical with other STs; CCUG15793 = CCUG18082, CCUG35214 = ST396, and S32F2 = ST35). Redundant sequences were removed, and phylogenetic analysis was performed on 444 unique sequences.

Neighbor-joining and minimum evolution analyses were conducted using MEGA4 (21) with distances corrected by the Kimura two-parameter method. Maximum likelihood analysis was performed using RAXML (Randomized Axelerated Maximum Likelihood) at the Cipres website (<http://8ball.sdsc.edu:8889/cipres-web/Home.do>). The bootstrap test was performed with 500 (neighbor joining and minimum evolution) or 100 (maximum likelihood) replicates.

Colony blot hybridization. The bacterial strains were dotted onto a gridded nitrocellulose membrane (Hybond-N⁺, RPN1782B; Amersham Biosciences) placed on a 9-cm-diameter chocolate agar plate and incubated at 35°C in a humidified atmosphere containing 5% CO₂ for 24 h. The study strains would form colonies of 0.5 to 2 mm under these conditions. The membranes with the bacterial colonies were carefully removed from the plates, and lysis of cells, plus binding and denaturation of DNA, was achieved by sequential placement of membranes on cellulose chromatography papers (Whatman 17 Chr) wetted with reagents in the following order: 10% SDS, 10 min; 0.5 M NaOH, 10 min; 1.5 M NaCl-0.5 M NaOH, 10 min; 1 M Tris-HCl, pH 7.4, 5 min; 1 M Tris-HCl, pH 7.4, 5 min; and 1.5 M NaCl-0.5 M Tris-HCl, pH 7.4, 10 min. After drying of membranes, cellular debris was wiped off with a paper towel. Membranes were then

preincubated with rotation for 1 h at 65°C in hybridization buffer (5 g/liter blocking reagent [Roche Diagnostics]–0.1% *N*-lauroylsarcosine–0.02% SDS–1× SSC [0.15 M NaCl plus 0.015 M sodium citrate]) and hybridized with digoxigenin-labeled probes (denatured by boiling for 10 min) in hybridization buffer for 1 h at 65°C. The membranes were then washed twice in 2× SSC–0.1% SDS at ambient temperature for 5 min and twice with 0.1× SSC–0.1% SDS at 65°C for 15 min. Detection of bound probe was assessed by alkaline phosphatase-labeled antidigoxigenin Fab fragments and a chromogenic assay (nitroblue tetrazolium–5-bromo-4-chloro-3-indolylphosphate) (Roche Diagnostics) according to the manufacturer's instructions. *Escherichia coli* strain MC1061 (29) was included on all membranes as a negative control.

Probes. Initial screening of strains was done by PCR, and the identity of amplified fragments from one strain was checked by sequencing. Subsequently, digoxigenin-labeled probes were generated from this strain by PCR using the PCR digoxigenin probe synthesis kit (Roche Diagnostics) according to the manufacturer's instructions. Two digoxigenin-labeled *hap* probes were amplified from *H. influenzae* strain Rd covering nt 1549 to 1963 (19) and nt 1549 to 2358 (using *hap* 2358r [GCTAGGCATTGTCCAAGTCGCATT] as reverse primer). *hia* was amplified from strain HK 1222 using primers *hia*-F and *hia*-R (8). DNA from strain HK 367^T was used as template in combination with the primers *hif*B-F and *hif*C-R (8) to amplify the *hif*BC intergenic region. *hmwC* was amplified with *hmwC*-F and *hmwC*-R (8), and *hmwA* was amplified with HMWP1HI and HMWP3RHI (43) using DNA from strain CCUG 35214 as template. *lic2B* 29f (CTTCAGCATATCAACGACGAGAGC) and *lic2B* 541r (TCGCGAGTATGATGGTGGATGTGA) were designed from the putative galactosyltransferase *Lic2B* gene (accession number AY091470) and used with DNA from strain HK367^T. *iga* probes 1, 3, and 4 were generated from strain HK 389^T by using primers PF1-PR1, LF-LR2, and CF1-CR1, respectively (44); *iga* probe 2 was amplified using primers *iga* 2063f (TAACAGGCGGAACAAACCTT)–*iga* 3060r (ATTGTTACTTGGTACGCTAG). With respect to strain Rd (gene *HI_0990*), *iga* probes 1 to 4 cover nt 480 to 1959, 2063 to 3060, 3088 to 4344, and 4246 to +26, respectively. *hemN* was amplified from strain CCUG 17210 by using primers *hemN* 219f and *hemN* 872r, and a *sodC* probe was amplified from *H. parainfluenzae* strain ATCC 29242 by using the primers from the work of Fung et al. (12). DNA from the type strain of *H. influenzae* (HK 389^T) was used as template in combination with the MLST primers (see above) to generate a *fucK* probe.

Nucleotide sequence accession numbers. DNA sequences generated in this study were deposited in GenBank under accession numbers EU909476 to EU909684.

RESULTS

The 42 study strains fermented D-glucose, were dependent on V factor, and were negative for β-galactosidase (*o*-nitrophenyl-β-D-galactopyranoside test) as expected for strains of the species *H. influenzae*, *H. aegyptius*, and *H. haemolyticus* (18).

Multilocus sequence phylogeny. The seven genes included in the *H. influenzae* MLST scheme were used for the analysis (26). The fuculokinase gene *fucK* was absent from 32 strains as indicated by the lack of a PCR product and confirmed by the lack of hybridization to a digoxigenin-labeled probe generated with the *fucK* primers and DNA from strain HK 389^T as template (not shown). The sequences from the six genes present in all strains (*adh*, *atpG*, *frdB*, *mdh*, *pgi*, and *recA*) were concatenated to a hybrid sequence of 2,712 nt and used for the cluster analysis. Figure 1A shows a dendrogram comparing the 42 study strains with the MLST STs available at the *H. influenzae* database by May 2007. The comparison is restricted to 2,712 nt, as the *fucK* sequences in the MLST STs were excluded. Where possible, the presentation in Fig. 1 follows the clustering described in the original MLST publication (26), dividing the species into phylogenetic groups I and II (blue and yellow, respectively). The position of each study strain or ST on the tree is listed in Table S2 in the supplemental material, together

with information on serotype and previous allocation to phylogenetic group (26) or clade (9).

Phylogenetic group I was composed of 343 STs and seven study strains (HK 389^T [type strain of *H. influenzae*], HK 1136 [serotype b], Rd, CCUG 35214, HK 1212, HK 367^T [type strain of *H. aegyptius*], and HK 1210). Phylogenetic group II encompassed 53 MLST STs and three study strains (HK 2067 [serotype f], HK 61, and HK 1212). Five STs (ST235, ST357, ST361, ST362, and ST364) and strain CCUG 30048 were located between the two main groups (gray in Fig. 1A); these sequences also segregate from phylogenetic groups I and II by minimum evolution and maximum likelihood analyses (not shown). Finally, three STs (ST35, ST363, and ST365) and 31 study strains were located on a common branch (32 unique sequences; red in Fig. 1A). For the purpose of this study, these 31 study strains will be referred to as variant strains. The three MLST STs located on the variant branch are throat isolates from healthy children (ST363 and ST365) and a cryptic genospecies biotype IV strain (ST35). Dendrograms constructed by minimum evolution and maximum likelihood were not identical with the neighbor-joining tree of Fig. 1A, but the cluster encompassing the variant strains was invariably present and supported by very high bootstrap values (neighbor joining, 99%; minimum evolution, 98%; maximum likelihood, 100%).

In Fig. 1B the phylogenetic analysis is restricted to the 42 study strains with the use of *H. parainfluenzae* strain T3T1 as the outgroup. The dendrogram reveals a tight cluster (Hi) encompassing the 11 strains depicted with strain designations in Fig. 1A and excluding the 31 variant strains. The remote position of strain 11p18 is primarily due to an atypical *frdB* sequence showing 97% similarity to the homologous gene in *H. parainfluenzae* strain T3T1. Some subgrouping is discernible in the variant cluster (red in Fig. 1B). The four cryptic genospecies biotype IV strains (11PS, 16N, S32F2, and 26E) are highly related, the mutual relationship between five mannose-fermenting variant strains of biotype VIII (CCUG 15794, CCUG 30047, CCUG 18082, CCUG 15793, and CCUG 36040) is supported by a significant bootstrap value of 85%, and a third cluster is formed by four porphyrin-synthesizing strains (CCUG 31732, CCUG 24145, CCUG 13929, and CCUG 32367). If the sequence analysis is reduced to a pairwise comparison with the type strain of *H. influenzae*, all strains of the Hi cluster are >96% identical to the type whereas 30 of 31 variant strains exhibit <96% identity to the type (values are given in Fig. 1B).

Delineation of Hi cluster in relation to number of genes analyzed. Figure 2 shows the discriminatory power of the cluster analysis in relation to the number of housekeeping genes included in the comparison. Neighbor-joining dendrograms were constructed from all combinations of gene fragments, and the bootstrap consensus tree was examined for the presence of a Hi cluster encompassing the same 11 strains as in Fig. 1B. The proportion of trees with identical Hi clusters, and the mean bootstrap support, was calculated. When the comparison was restricted to single genes, only the analysis based on *pgi* produced the Hi cluster, supported by an untenable bootstrap value of 18%. Twelve of 20 possible three-gene combinations separated Hi clusters from variant strains, supported by a mean bootstrap value of 68% (range, 39 to 90%). With the present study material, five genes were needed for a robust

FIG. 1. Neighbor-joining dendrograms based on concatenated sequences of housekeeping genes *adh*, *atpG*, *frdB*, *mdh*, *pgi*, and *recA* (2,712 nt). Phylogenetic group I, blue; phylogenetic group II, yellow; variant cluster, red. (A) Forty-two study strains on a background of 447 *H. influenzae* STs from the MLST database. (B) Forty-two study strains rooted by *H. parainfluenzae* strain T3T1. Values to the left of nodes are percentages of bootstrap replications supporting the node. Selected phenotypic and genotypic characters are given. Bt, *H. influenzae* biotype; Po, synthesis of porphyrins from δ -aminolevulinic acid; He, hemolysis; Ga, gas from glucose; As, asaccharolytic with respect to fermentation of galactose, ribose, and xylose; Su, fermentation of sucrose; Mn, fermentation of D-mannose; Ig, cleavage of IgA1; H₂S, emission of hydrogen sulfide; ha, *hap*; hi, *hia*; hf, *hif*; hm, *hmv*; fK, *fucK*; ig, *iga*; lb, *lic2B*; sC, *sodC*; Sim, nucleotide identity in comparison with the corresponding gene fragments in the type strain of *H. influenzae* (2,712 nt). Bars represent 0.5 and 2% residue substitutions (A and B, respectively).

core plus the f branch was supported by a bootstrap value of 91%. If both the f branch and the e branch were included (i.e., excluding all variant strains), the bootstrap support dropped to 57%. The four cryptic genospecies biotype IV strains were located at the most distant end of the unrooted dendrogram (IV in Fig. 3B).

Phenotypic traits. Phenotypic profiling of the study strains revealed a number of phenotypic characters with preponderance in either Hi cluster strains or variant strains (Fig. 1B). Production of gas by fermentation of glucose was observed with seven variant strains and not with strains of the Hi cluster. Production of H_2S was assessed by discoloration (after 48 h) of lead acetate paper placed in the lid of an inoculated chocolate agar plate. Many strains gave clear-cut positive or negative results, but a weak discoloration of the paper occurred repeatedly with some strains, including the type strain of *H. influenzae*. When the weakly positive strains were scored as negative, 10 of 11 Hi cluster strains were negative for production of H_2S , whereas 26 of 31 variant strains were positive.

Seven variant strains were hemolytic on horse blood agar. Three hemolytic strains fermented D-mannose; a fourth strain was not dependent on X factor. The type strain of *H. haemolyticus* was markedly saccharolytic, capable of fermenting D-maltose, maltotriose, and D-mannose, in addition to D-glucose and D-ribose (not shown).

Twelve variant strains fermented D-mannose, a property which was not observed with Hi cluster strains.

Eleven variant strains synthesized porphyrins from δ -amino-levulinic acid and were independent of X factor. These strains were all capable of fermenting sucrose, in contrast to the rest of the study material. In addition to D-glucose and sucrose, further saccharolytic activity was detected with only four strains, fermenting D-mannose, D-malate, D-galactose, and D-galactose plus maltotriose, respectively. All X-factor-independent strains were positive for urease activity; one was hemolytic.

IgA1 protease activity and detection of *iga*. The absence of *iga*, the gene encoding the IgA1-specific serine endopeptidase, has been used to define *H. haemolyticus* (18, 25, 30, 46). When the present study strains were assessed for IgA1 protease activity, 10 of 11 strains of the Hi cluster but also 13 strains excluded from this cluster were positive (Fig. 1B). We also tested for the presence of *iga* by hybridization assays. Probes were generated spanning four different sections of the gene and using DNA from the type strain of *H. influenzae* as template (probe sequences were deposited in GenBank under accession number EU909681). The locations of the probes in the *iga* gene of strain Rd and the hybridization signals of individual probes with the study strains are shown in Fig. 4.

Clear and identical results were obtained with probes hybridizing with the linker region (probe 3) and the β -core domain (probe 4) of *iga*, which were positive for all strains of the Hi cluster, including the one strain without activity, and for 20 of 31 variant strains. IgA1 protease activity was undetectable with eight study strains testing positive for the *iga* gene by these two probes. Probe 2, generated from the region of *iga* that encodes the carboxyl-terminal part of the mature protease, showed similar reactivity with only a single discrepant result: strain CCUG 30047 (dot 19) was positive with probes 3 and 4 but negative with probe 2. Probe 1, spanning the middle part of

FIG. 4. Detection of IgA-specific serine endopeptidase gene *iga* by colony blot hybridization assay. Top, location of hybridization probes with reference to gene HI_0990 of strain Rd. Bottom, hybridization results using DNA from the type strain of *H. influenzae* (HK 389^T) as template for generation of hybridization probes. Study strains were dotted onto the membranes in alphabetical order from top left (dot numbers are shown for the top row, for the first and last dots on the following lines, and for dot 42; numbering of strains is given in Table S1 in the supplemental material). E, *Escherichia coli* strain MC1061 (negative control).

the mature protease, hybridized differently with several additional strains testing negative (Fig. 4). The results obtained with probes 3 and 4 are included in Fig. 1B; these probes correspond closely to probes used in previous investigations (25, 30, 46).

The linker region of the *iga* gene varies in length between different strains of *H. influenzae* (38, 44). We also generated *iga* linker region probes using strain Rd and the encapsulated type b strain HK 368 as templates. Although shorter probes were generated from these strains (Rd, 1,257 nt; HK 368, 813 nt) due to deletions relative to the *iga* gene in HK 389^T, identical hybridization patterns with the study strains were observed (not shown).

Prevalence of other virulence and marker genes. Clinical isolates of *H. influenzae* encode different combinations of adhesins associated with encapsulation and anatomical site of isolation, which may be decisive for differences in host cell binding specificities (40, 43). We searched for the *H. influenzae* adherence and penetration protein gene *hap* (15), the *H. influenzae* adhesion protein gene *hia* (2), the high-molecular-weight adhesion genes *hmw1A* and *hmwC* (43), the pilus gene *hif* cluster (28), and the lipooligosaccharide biosynthesis gene *lic2B* (36).

The results of the probe hybridization assays are given in Fig. 1B. In accordance with previous data (9), *hia* and *lic2B* were consistently present in phylogenetic group II strains and variably present in group I strains. The pilus gene cluster and *lic2B* were sporadically detected in variant strains, while *hia* and *hmw* were not. Based on a limited number of isolates, these putative virulence genes were variably but more regularly detected in Hi cluster strains.

The presence of *hap* was significantly associated with the Hi cluster. All strains of the Hi cluster and only a single variant strain (HK 855) hybridized with a 415-nt probe of *hap* generated from strain Rd. Other genes of possible value in the delineation of *H. influenzae* were also examined. *fucK*, which encodes fuculokinase and is part of the MLST scheme, was a promising marker, being absent from all variant strains and

from only 1 (CCUG 30048) of 11 Hi cluster strains. The Cu,Zn-superoxide dismutase gene *sodC*, which has been used as a marker of cryptic genospecies biotype IV (22) and *H. haemolyticus* (12), was detected with all variant strains and with only one strain of the Hi cluster (HK 2067 of serotype f) (Fig. 2).

Hemin biosynthesis. Genes involved in hemin biosynthesis were searched for by use of degenerate PCR primers designed to hybridize with conserved regions within the respective genes (see Materials and Methods), and *hemB* (prophobilinogen synthase), *hemE* (uroporphyrinogen decarboxylase), and *hemN* (coproporphyrinogen dehydrogenase) were identified with the respective primer sets. The sequences from strain CCUG 30218 have been deposited in GenBank (accession numbers EU909682 to EU909684). The putative *hemB* gene encodes a polypeptide of 340 amino acids showing 92% amino acid identity to the corresponding enzymes from both *H. parainfluenzae* and *A. actinomycetemcomitans*. The gene is located between the translocase protein gene *tatC* and the glutamate dehydrogenase gene *gdhA*, similar to the position in the *H. parainfluenzae* T3T1 genome. The putative *hemE* gene of strain CCUG 30218 encodes a polypeptide of 354 amino acids showing 90, 85, and 82% amino acid identity to the corresponding enzymes from *H. parainfluenzae*, *P. multocida*, and *A. actinomycetemcomitans*, respectively. The gene is located between an NAD pyrophosphatase gene (*nudC*) and a hypothetical conserved gene (Clusters of Orthologous Groups of proteins 3068), similar to the position in the *A. actinomycetemcomitans*, *H. parainfluenzae*, *Histophilus somni*, and *P. multocida* genomes. The 568-nt fragment of the putative *hemN* gene showed 77% nucleotide sequence similarity to the corresponding genes from both *H. parainfluenzae* T3T1 and *P. multocida* Pm70. Flanking genes of *hemN* in CCUG 30218 were not identified. A digoxigenin-labeled 654-nt PCR amplicon from strain CCUG 30218 hybridized with DNA from all 11 X-factor-independent strains and not with DNA from any strain unable to synthesize porphyrins (not shown).

DISCUSSION

The genetic diversity and population structure of *H. influenzae* have previously been addressed (9, 26, 32, 41). All such studies were based on collections of strains that possessed a pattern of phenotypic traits generally expected to be able to define the species. A bipartite division was detected based on multilocus enzyme electrophoresis (32) and modified by MLST with designations of phylogenetic groups I and II (26). Based on maximum-parsimony analysis of 359 STs defined by MLST analysis, a somewhat different population structure was recently suggested (9). The study identified 13 clades containing 6 to 89 STs, whereas 80 STs were not included in any of the clades. Clade 2 corresponds closely to phylogenetic group II, and the major modification relates to the dissection of phylogenetic group I into the remaining 12 clades and STs outside clades (9).

In the present study we included a number of closely related but supposedly distinct taxa in the comparison to test the cohesiveness of the species *H. influenzae* and its separation from neighboring taxa. The resulting population structure was tripartite. A variant cluster encompassing 31 study strains and

three STs from the MLST database was invariably present and supported by high bootstrap values. The delineation between *H. influenzae* phylogenetic groups I and II was more uncertain, and depending on the computational algorithm, a number of strains were intermediately positioned on the trees. Three STs, ST361, ST362, and ST364, were consistently located outside the phylogenetic groups; together with ST363 and ST365 of the variant cluster, these STs represent throat isolates recovered from healthy children in Michigan.

Although the existence of a variant cluster was strongly supported by statistical resampling procedures, it was dependent on a sufficient number of gene fragments being included in the analysis. We obtained a satisfying delineation of a Hi cluster by phylogenetic analysis of five or six housekeeping genes, whereas comparison of ≤ 4 gene fragments could be misleading. Hi cluster strains were also related by 16S rRNA gene sequence, although in this case, a clear demarcation from variant strains was not observed. Compared with housekeeping gene phylogeny based on multiple loci, the resolving power of 16S rRNA gene sequence comparison is hampered by shorter sequences and less variation.

Support for delineation between the Hi cluster and variant strains was also provided by other genotypic and phenotypic traits. Strains with the presence of hemolysis, synthesis of porphyrins, or fermentation of mannose qualified for inclusion in the study as variant strains, and phenotypic profiling showed emission of H₂S; production of gas from glucose; and inability to ferment galactose, ribose, and xylose to be preferentially associated with strains excluded from the Hi cluster. Detection of H₂S was the single phenotypic character most strongly correlated with the phylogenetic distinction, with 10 of 11 Hi cluster strains being negative and 26 of 31 variant strains being positive. For broader use, the test for H₂S should be optimized and standardized in order to remedy the subjective nature of the assessment of the discoloration of the lead acetate paper. Detection of IgA1 protease activity was a less reliable marker of the Hi cluster, as more than one-third of the variant strains, including strains assigned to *H. haemolyticus*, were able to cleave IgA1 (Fig. 1B). Surprisingly, the *iga* gene, encoding the IgA1-specific serine endopeptidase, was detected in eight additional variant strains negative for IgA1 protease activity. The molecular mechanism of the lack of expression of IgA1 protease in these strains was not elucidated, but our hybridization results clearly do not support the adduced use of *iga* as a decisive marker of *H. influenzae* (25, 30, 46).

The fuculokinase gene *fucK*, the *H. influenzae* adherence and penetration protein gene *hap*, and the Cu,Zn-superoxide dismutase gene *sodC* were differentially present in the study strains and may be valuable in the discrimination of the Hi cluster from variant strains. *fucK* was absent from all strains excluded from the Hi cluster and from a single strain included in this cluster (see below). We detected *hap* in all strains of the Hi cluster and in one (strain HK 855) of 31 variant strains using a previously described 415-nt probe hybridizing with a conserved region in the gene (19). When a larger probe (810 nt) was employed, an additional strain (CCUG 50565) was positive, and homologs of *hap* or remnants of the gene may be present in a larger fraction of strains. *hap* has previously been detected in 53 of 53 non-type b encapsulated strains of *H. influenzae*, whereas the gene product was expressed in only 28

of the strains (40). A conserved deletion leading to pseudogene formation has been described for a particular subset of *H. influenzae* strains (*H. aegyptius* and Brazilian purpuric fever-associated strains of *H. influenzae* biogroup *aegyptius*) (19). Hybridization for dispensable virulence genes is subject to uncertainty, but irrespective of the possible degeneracy or expression of *hap*, large differences in hybridization signal were observed between Hi cluster strains and variant strains with the present selection of probes.

Detection of *sodC* has been used to distinguish cryptic genospecies biotype IV (22) and *H. haemolyticus* (12) from *H. influenzae*. We detected *sodC* in 31 of 31 variant strains and in 1 (HK 2067 of serotype f) of 11 *H. influenzae* strains. *sodC* is located adjacent to *bexA* in encapsulated strains of phylogenetic group II (42, 47) and has hitherto not been described in nonencapsulated strains.

A crucial question is the possible synonymy of the Hi cluster with *H. influenzae*. Two strains received as "*Haemophilus intermedius*," CCUG 30048 and CCUG 35214, were included in the Hi cluster. Based on the full seven-gene MLST scheme, strain CCUG 35214 was allocated to the previously described ST396, and 16S rRNA gene sequence identity to the type strain of *H. influenzae* exceeded 99.8%. By phenotype, CCUG 35214 was a typical representative of *H. influenzae* with the sole exception that the strain was negative for IgA1 protease activity. The strain was positive for the *iga* gene by hybridization (Fig. 4, dot 24), and failure to express the protease cannot challenge the specific allocation so clearly documented by the genotype and by other phenotypic means. Strain CCUG 30048 was most closely related to ST235 and ST357 by multilocus sequence phylogeny, although none of the six allelic sequences have been previously deposited in the MLST database. It possesses the putative *H. influenzae* virulence genes *hap*, *hia*, and *hif* and corresponds by phenotype to *H. influenzae* biotype IV, except for the detectable emission of hydrogen sulfide. By 16S rRNA gene sequence comparison, it takes an intermediate position (Fig. 3A), and most conspicuously, it is negative for *fucK* by both PCR and hybridization. *fucK* is part of the MLST scheme for typing of *H. influenzae*, and strains of *H. influenzae* negative for *fucK* have not been reported before; however, no biological reasons rule out the existence of such strains. The *fucK*⁺ genotype is not exclusively associated with *H. influenzae*, as several strains of the variant cluster have been allocated to MLST STs after being successfully typed with the full seven-gene MLST scheme (ST 35, ST 363, and ST365). Variant strains were included in this study due to aberrant phenotypic or genetic characters. Strain CCUG 30048 may represent a borderline type, and such strains may go unnoticed unless an extensive characterization is carried out. It is likely that other such strains could be identified, and inclusion of a larger number of borderline strains could impede the delineation based on multilocus sequence phylogeny. The present demarcation of the Hi cluster may spring from the selection of study strains and does not negate the possibility of a continuum of genotypes from the core of the species and outwards.

Another problem with the delineation of *H. influenzae* is the definition of neighboring species. Two validly published species are closely related to *H. influenzae* by both phenotypic and genotypic means, namely, *H. aegyptius* and *H. haemolyticus*. The type strain of *H. aegyptius* (HK 367^T) is located within the

Hi cluster by multilocus sequence phylogeny and by 16S rRNA gene sequence comparison, in line with the previous demonstration that the close DNA relationship of *H. influenzae* and *H. aegyptius* is in conflict with their designation as separate species (5). Six study strains corresponded by phenotype to *H. haemolyticus* (18). These hemolytic strains were dependent on X and V factors; negative for ornithine decarboxylase, xylose, and galactose fermentation; positive for urease; and variable with respect to production of indole and fermentation of ribose and mannose. However, by sequence analysis, several of these hemolytic strains did not cluster with the type strain of *H. haemolyticus*, irrespective of whether the comparison was based on housekeeping genes or 16S rRNA gene sequences (Fig. 1B and 3A). Accordingly, the species *H. haemolyticus*, as presently defined, is polyphyletic. The sequential isolation of hemolytic and nonhemolytic strains of identical pulsed-field electrophoresis genotypes from a healthy adult carrier has been demonstrated (30), and the property is sometimes lost in the laboratory by subculture (17). This implies that the hemolytic phenotype is unstable and that the character is of minor taxonomic importance. The definition of *H. haemolyticus* could therefore be broadened to encompass a wider spectrum of strains, as has been done in a number of recent investigations (25, 30, 31). The variant strains of the present study formed a coherent sequence cluster based on the investigated gene fragments; the mean pairwise sequence distance between 30 variant strains (excluding strain 11p18) was 3.4% (maximal divergence, 4.9%), which is comparable to the mean pairwise distance between 11 Hi cluster strains (3.3%; maximal divergence, 5.1%). To classify the 31 variant strains of this study with *H. haemolyticus* could be a convenient description since *H. haemolyticus* is a validated specific epithet in proximity to, but distinct from, *H. influenzae*. However, such grouping of strains conflicts with current definitions of bacterial species based on phenotypic traits and would necessitate further substantiation, preferably from whole-genome-based sequence comparisons.

Based on DNA hybridization data, Burbach (4) described strains related to *H. influenzae* but capable of synthesizing porphyrins and therefore not dependent on X factor, in conflict with the traditional definition of the species (16). This finding has not received much attention, probably related to the rare occurrence and difficult identification of such strains. The three strains originally identified by DNA hybridization were phenotypically characterized by fermentation of sucrose (another property in conflict with the traditional definition of *H. influenzae*) and negative tests for ornithine decarboxylase, β -galactosidase, and maltose fermentation, plus a weak fermentation of fructose. According to this phenotype, seven additional strains of "*H. intermedius* subsp. *intermedius*" have been identified by the CCUG (www.ccug.se). Our own isolate (PN24) was positive for ornithine decarboxylase, and this may indicate that the phenotype described by Burbach based on a limited number of strains could be incidental. Eleven strains were available for this study, and the synthesis of porphyrins and lack of dependence on X factor were confirmed. Identification and sequencing of hemin biosynthesis genes *hemB*, *hemE*, and *hemN* showed low nucleotide similarities to similar genes previously deposited in the databases, a finding which is contradictory to a recent transfer of these genes from, e.g., *H.*

parainfluenzae. Flanking genes of *hemB* and *hemE* were identified and found to be comparable with gene orders in related bacteria. Our data suggest that hemin biosynthesis in these variant strains is chromosomally encoded and could represent an ancestral genotype, from which the X-factor-dependent *H. influenzae* evolved.

The results of this study exemplify the problems with current definitions of bacterial species based on phenotypic traits. Although strains previously assigned to *H. haemolyticus* were part of a deeply branching sequence cluster distinct from traditional *H. influenzae*, the same cluster included phenotypically different strains of “*H. intermedius*” and cryptic genospecies biotype IV. Although *H. influenzae* is phenotypically more homogeneous than other *Haemophilus* species, the genetic diversity and multicluster structure of strains traditionally associated with *H. influenzae* make it difficult to define natural borders of that species.

ACKNOWLEDGMENTS

Flemming Scheutz, WHO International Escherichia and Klebsiella Centre, Statens Seruminstitut, Copenhagen, Denmark, is thanked for helpful suggestions concerning colony blot assay, and student programmer Poul Liboriussen, Bioinformatics Research Center, University of Aarhus, is thanked for assistance with RAXML computations.

This work was supported by a grant from the Danish Lung Association.

REFERENCES

- Albritton, W. L., J. L. Brunton, M. Meier, M. N. Bowman, and L. A. Slaney. 1982. *Haemophilus influenzae*: comparison of respiratory tract isolates with genitourinary tract isolates. *J. Clin. Microbiol.* **16**:826–831.
- Barenkamp, S. J., and J. W. St. Geme. 1996. Identification of a second family of high-molecular-weight adhesion proteins expressed by non-typable *Haemophilus influenzae*. *Mol. Microbiol.* **19**:1215–1223.
- Brenner, D. J., L. W. Mayer, G. M. Carlone, L. H. Harrison, W. F. Bibb, M. C. D. Brandileone, F. O. Sotnek, K. Irino, M. W. Reeves, J. M. Swenson, K. A. Birkness, R. S. Weyant, S. F. Berkley, T. C. Woods, A. G. Steigerwalt, P. A. D. Grimont, R. M. McKinney, D. W. Fleming, L. L. Gheesling, R. C. Cooksey, R. J. Arko, C. V. Broome, and the Brazilian Purpuric Fever Study Group. 1988. Biochemical, genetic, and epidemiologic characterization of *Haemophilus influenzae* biogroup aegyptius (*Haemophilus aegyptius*) strains associated with Brazilian purpuric fever. *J. Clin. Microbiol.* **26**:1524–1534.
- Burbach, S. 1987. Reklassifizierung der Gattung *Haemophilus* Winslow et al. 1917 auf Grund der DNA-Basensequenzhomologie. Doctoral thesis. Philipps-Universität Marburg, Marburg, Germany.
- Casin, I., F. Grimont, and P. A. Grimont. 1986. Deoxyribonucleic acid relatedness between *Haemophilus aegyptius* and *Haemophilus influenzae*. *Ann. Inst. Pasteur Microbiol.* **137B**:155–163.
- Challacombe, J. F., A. J. Duncan, T. S. Brettin, D. Bruce, O. Chertkov, J. C. Deter, C. S. Han, M. Misra, P. Richardson, R. Tapia, N. Thayer, G. Xie, and T. J. Inzana. 2007. Complete genome sequence of *Haemophilus somnus* (*Histophilus somni*) strain 129Pt and comparison to *Haemophilus ducreyi* 35000HP and *Haemophilus influenzae* Rd. *J. Bacteriol.* **189**:1890–1898.
- Cohan, F. M. 2002. What are bacterial species? *Annu. Rev. Microbiol.* **56**:457–487.
- Ecevit, Z., K. W. McCrea, M. M. Pettigrew, A. Sen, C. F. Marrs, and J. R. Gilsdorf. 2004. Prevalence of the *hifBC*, *hmv1A*, *hmv2A*, *hmvC*, and *hia* genes in *Haemophilus influenzae* isolates. *J. Clin. Microbiol.* **42**:3065–3072.
- Erwin, A. L., S. A. Sandstedt, P. J. Bonthuis, J. L. Geelhood, K. L. Nelson, W. C. Unrath, M. A. Diggle, M. J. Theodore, C. R. Pleatman, E. A. Mothershead, C. T. Sacchi, L. W. Mayer, J. R. Gilsdorf, and A. L. Smith. 2008. Analysis of genetic relatedness of *Haemophilus influenzae* isolates by multilocus sequence typing. *J. Bacteriol.* **190**:1473–1483.
- Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, T. F. Tomb, B. A. Dougherty, J. M. Merrick, et al. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**:496–512.
- Fraser, C., W. P. Hanage, and B. G. Spratt. 2007. Recombination and the nature of bacterial speciation. *Science* **315**:476–480.
- Fung, W. W., C. A. O'Dwyer, S. Sinha, A. L. Brauer, T. F. Murphy, J. S. Kroll, and P. R. Langford. 2006. Presence of copper- and zinc-containing superoxide dismutase in commensal *Haemophilus haemolyticus* isolates can be used as a marker to discriminate them from nontypeable *H. influenzae*. *J. Clin. Microbiol.* **44**:4222–4226.
- Gousset, N., A. Rosenau, P. Y. Sizaret, and R. Quentin. 1999. Nucleotide sequences of genes coding for fimbrial proteins in a cryptic genospecies of *Haemophilus* spp. isolated from neonatal and genital tract infections. *Infect. Immun.* **67**:8–15.
- Hedegaard, J., H. Okkels, B. Bruun, M. Kilian, K. K. Mortensen, and N. Nørskov-Lauritsen. 2001. Phylogeny of the genus *Haemophilus* as determined by comparison of partial *infB* sequences. *Microbiology* **147**:2599–2609.
- Hendrixson, D. R., and J. W. St. Geme. 1998. The *Haemophilus influenzae* Hap serine protease promotes adherence and microcolony formation, potentiated by a soluble host protein. *Mol. Cell* **2**:841–850.
- Kilian, M. 1974. A rapid method for the differentiation of *Haemophilus* strains. The porphyrin test. *Acta Pathol. Microbiol. Scand. Sect. B* **82**:835–842.
- Kilian, M. 1976. A taxonomic study of the genus *Haemophilus*, with the proposal of a new species. *J. Gen. Microbiol.* **93**:9–62.
- Kilian, M. 2005. Genus III *Haemophilus* Winslow, Broadhurst, Buchanan, Rogers and Smith 1917, 561^{AL}, p. 883–904. In D. J. Brenner, N. R. Krieg, J. T. Staley, and G. M. Garrity (ed.), *Bergey's manual of systematic bacteriology*, 2nd ed., vol. 2, part B. Springer, New York, NY.
- Kilian, M., K. Poulsen, and H. Lomholt. 2002. Evolution of the paralogous *hap* and *iga* genes in *Haemophilus influenzae*: evidence for a conserved hap pseudogene associated with microcolony formation in the recently diverged *Haemophilus aegyptius* and *H. influenzae* biogroup aegyptius. *Mol. Microbiol.* **46**:1367–1380.
- Kuhnert, P., J. Frey, N. P. Lang, and L. Mayfield. 2002. Phylogenetic analysis of *Prevotella nigrescens*, *Prevotella intermedia* and *Porphyromonas gingivalis* clinical strains reveals a clear species clustering. *Int. J. Syst. Evol. Microbiol.* **52**:1391–1395.
- Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* **5**:150–163.
- Langford, P. R., B. J. Sheehan, T. Shaikh, and J. S. Kroll. 2002. Active copper- and zinc-containing superoxide dismutase in the cryptic genospecies of *Haemophilus* causing urogenital and neonatal infections discriminates them from *Haemophilus influenzae* sensu stricto. *J. Clin. Microbiol.* **40**:268–270.
- Mak, G. C., P. L. Ho, C. W. Tse, S. K. Lau, and S. S. Wong. 2005. Reduced levofloxacin susceptibility and tetracycline resistance in a clinical isolate of *Haemophilus quentini* identified by 16S rRNA sequencing. *J. Clin. Microbiol.* **43**:5391–5392.
- May, B. J., Q. Zhang, L. L. Li, M. L. Paustian, T. S. Whittam, and V. Kapur. 2001. Complete genomic sequence of *Pasteurella multocida* Pm70. *Proc. Natl. Acad. Sci. USA* **98**:3460–3465.
- McCrea, K. W., J. Xie, N. LaCross, M. Patel, D. Mukundan, T. F. Murphy, C. F. Marrs, and J. R. Gilsdorf. 2008. Relationships of nontypeable *Haemophilus influenzae* strains to hemolytic and nonhemolytic *Haemophilus haemolyticus* strains. *J. Clin. Microbiol.* **46**:406–416.
- Meats, E., E. J. Feil, S. Stringer, A. J. Cody, R. Goldstein, J. S. Kroll, T. Popovic, and B. G. Spratt. 2003. Characterization of encapsulated and non-encapsulated *Haemophilus influenzae* and determination of phylogenetic relationships by multilocus sequence typing. *J. Clin. Microbiol.* **41**:1623–1636.
- Mestecky, J., and M. Kilian. 1985. Immunoglobulin A (IgA). *Methods Enzymol.* **116**:37–75.
- Mhlanga-Mutangadura, T., G. Morlin, A. L. Smith, A. Eisenstark, and M. Golomb. 1998. Evolution of the major pilus gene cluster of *Haemophilus influenzae*. *J. Bacteriol.* **180**:4693–4703.
- Minton, N. P. 1984. Improved plasmid vectors for the isolation of translational *lac* gene fusions. *Gene* **31**:269–273.
- Mukundan, D., Z. Ecevit, M. Patel, C. F. Marrs, and J. R. Gilsdorf. 2007. Pharyngeal colonization dynamics of *Haemophilus influenzae* and *Haemophilus haemolyticus* in healthy adult carriers. *J. Clin. Microbiol.* **45**:3207–3217.
- Murphy, T. F., A. L. Brauer, S. Sethi, M. Kilian, X. Cai, and A. J. Lesse. 2007. *Haemophilus haemolyticus*: a human respiratory tract commensal to be distinguished from *Haemophilus influenzae*. *J. Infect. Dis.* **195**:81–89.
- Musser, J. M., J. S. Kroll, E. R. Moxon, and R. K. Selander. 1988. Evolutionary genetics of the encapsulated strains of *Haemophilus influenzae*. *Proc. Natl. Acad. Sci. USA* **85**:7758–7762.
- Nørskov-Lauritsen, N., B. Bruun, and M. Kilian. 2005. Multilocus sequence phylogenetic study of the genus *Haemophilus* with description of *Haemophilus pittmaniae* sp. nov. *Int. J. Syst. Evol. Microbiol.* **55**:449–456.
- Nørskov-Lauritsen, N., and M. Kilian. 2006. Reclassification of *Actinobacillus actinomycetemcomitans*, *Haemophilus aphrophilus*, *Haemophilus paraphrophilus* and *Haemophilus segnis* as *Aggregatibacter actinomycetemcomitans* gen. nov., comb. nov., *Aggregatibacter aphrophilus* comb. nov. and *Aggregatibacter segnis* comb. nov., and emended description of *Aggregatibacter aphrophilus* to include V factor-dependent and V factor-independent isolates. *Int. J. Syst. Evol. Microbiol.* **56**:2135–2146.
- Olsen, I., F. E. Dewhirst, B. J. Paster, and H.-J. Busse. 2005. Family I. *Pasteurellaceae* Pohl 1981, 382^{VP}, p. 851–856. In D. J. Brenner, N. R. Krieg, J. T. Staley, and G. M. Garrity (ed.), *Bergey's manual of systematic bacteriology*, 2nd ed., vol. 2, part B. Springer, New York, NY.

36. Pettigrew, M. M., B. Foxman, C. F. Marrs, and J. R. Gilsdorf. 2002. Identification of the lipooligosaccharide biosynthesis gene *lic2B* as a putative virulence factor in strains of nontypeable *Haemophilus influenzae* that cause otitis media. *Infect. Immun.* **70**:3551–3556.
37. Pittman, M., and D. J. Davis. 1950. Identification of the Koch-Weeks bacillus (*Haemophilus aegyptius*). *J. Bacteriol.* **59**:413–426.
38. Poulsen, K., J. Reinholdt, and M. Kilian. 1992. A comparative genetic study of serologically distinct *Haemophilus influenzae* type 1 immunoglobulin A1 proteases. *J. Bacteriol.* **174**:2913–2921.
39. Quentin, R., R. Ruimy, A. Rosenau, J. M. Musser, and R. Christen. 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**:1380–1385.
40. Rodriguez, C. A., V. Avadhanula, A. Buscher, A. L. Smith, J. W. St. Geme, and E. E. Adderson. 2003. Prevalence and distribution of adhesins in invasive non-type b encapsulated *Haemophilus influenzae*. *Infect. Immun.* **71**:1635–1642.
41. Sacchi, C. T., D. Alber, P. Dull, E. A. Mothershed, A. M. Whitney, G. A. Barnett, T. Popovic, and L. W. Mayer. 2005. High level of sequence diversity in the 16S rRNA genes of *Haemophilus influenzae* isolates is useful for molecular subtyping. *J. Clin. Microbiol.* **43**:3734–3742.
42. Satola, S. W., P. L. Schirmer, and M. M. Farley. 2003. Genetic analysis of the capsule locus of *Haemophilus influenzae* serotype f. *Infect. Immun.* **71**:7202–7207.
43. van Schilfgaarde, M., P. van Ulsen, P. Eijk, M. Brand, M. Stam, J. Kouame, L. van Alphen, and J. Dankert. 2000. Characterization of adherence of nontypeable *Haemophilus influenzae* to human epithelial cells. *Infect. Immun.* **68**:4658–4665.
44. Vitovski, S., K. T. Dunkin, A. J. Howard, and J. R. Sayers. 2002. Nontypeable *Haemophilus influenzae* in carriage and disease. *JAMA* **287**:1699–1705.
45. Reference deleted.
46. Xie, J., P. C. Juliao, J. R. Gilsdorf, D. Ghosh, M. Patel, and C. F. Marrs. 2006. Identification of new genetic regions more prevalent in nontypeable *Haemophilus influenzae* otitis media strains than in throat strains. *J. Clin. Microbiol.* **44**:4316–4325.
47. Zhou, J., D. K. Law, M. L. Sill, and R. S. Tsang. 2007. Nucleotide sequence diversity of the *bexA* gene in serotypeable *Haemophilus influenzae* strains recovered from invasive disease patients in Canada. *J. Clin. Microbiol.* **45**:1996–1999.