Published in final edited form as:
*J Microbiol Methods*. 2008 October ; 75(2): 369–371. doi:10.1016/j.mimet.2008.06.023.

# Comparison of laboratory-based and phylogenetic methods to distinguish between *Haemophilus influenzae* and *H. haemolyticus*

**Sara A. Sandstedt**[1,*], **Lixin Zhang**[1], **Mayurika Patel**[2], **Kirk W. McCrea**[2], **Zhaohui Qin**[3], **Carl F. Marrs**[1], and **Janet R. Gilsdorf**[1,2]

1*Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, Michigan*

2*Department of Pediatrics and Communicable Diseases, University of Michigan Medical School, Ann Arbor, Michigan*

3*Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan*

## Summary

New methods to distinguish between nontypeable *Haemophilus influenzae* and nonhemolytic *Haemophilus haemolyticus* were compared. The results of *iga* variable region hybridization to dotblots and library-on-a-slide microarrays were more similar to a “gold standard” multigene phylogenetic tree than *iga* conserved region hybridization or P6 7F3 epitope immunoblots.

### Keywords

*Haemophilus haemolyticus*; Nontypeable *Haemophilus influenza*; *Haemophilus* classification; P6 7F3 epitope; immunoglobulin A1 protease

Nontypeable *Haemophilus influenzae* (NTHi) are gram-negative coccobacilli that asymptomatically colonize the nasopharynx, but cause respiratory infections such as bronchitis, otitis media, and sinusitis. Standard tests to identify *H. influenzae* in clinical specimens, X- and V- factor dependence (Kilian, 2005), do not distinguish between *H. influenzae* and a related, non-pathogenic species, *Haemophilus haemolyticus*. Such distinction has relied on the ability of *H. haemolyticus* to lyse horse red blood cells, which *H. influenzae* cannot do. However, this phenotype may be lost during passage (Kilian, 1976). Worse, it was reported recently that some strains of even low passage *H. haemolyticus* do not hemolyze, making this test unreliable (Murphy et al., 2007). In fact, between 4% and 27% of throat culture strains originally designated *H. influenzae* appear to be *H. haemolyticus* (Juliao et al., 2007, Mukundan et al., 2007, Murphy et al., 2007), a concern for research using throat strains. Because *H. haemolyticus* does not cause disease or live in normally sterile sites (Murphy et al., 2007, Xie et al., 2006), most clinical samples do not need to be tested for the presence of *H. haemolyticus* and may be tested for *H. influenzae* in the traditional manner, by X and V factor dependence (McCrea et al., 2008).

*Corresponding author: Sara A. Sandstedt, University of Michigan, 1420 Washington Heights, Department of Epidemiology, SPH building II, Ann Arbor, MI 48109-2029, Phone: (734)615-4866, Fax: (734)936-6732, Email: sandsted@umich.edu.

To better distinguish between *H. influenzae* and *H. haemolyticus* in studies of *H. influenzae* pharyngeal colonization, McCrea et al. (2008) used 88 strains of *H. influenzae*, 33 strains of hemolytic *H. haemolyticus*, 76 strains of nonhemolytic *H. haemolyticus* and five outgroup taxa to reconstruct a minimum evolution (ME) phylogenetic tree based on the concatenated sequences of five genes: *adk*, *pgi*, *recA*, *infB*, and 16S rRNA. Known *H. influenzae* and *H. haemolyticus* are found in separate clusters and biochemical tests correlate well with the tree structure. While such a tree serves as a "gold standard" to distinguish *H. influenzae* from nonhemolytic *H. haemolyticus*, classifying new strains by placing them in the tree involves sequencing portions of five genes and significant computational time. Trees reconstructed using single genes were not, by themselves, able to separate the two species.

Rapid methods for research laboratories to distinguish between *H. influenzae* and nonhemolytic *H. haemolyticus* have been suggested. The *iga* gene encoding the enzyme immunoglobulin A1 protease (Kilian, 2005) and the 7F3 epitope of P6 are present in *H. influenzae* but not *H. haemolyticus* (Murphy et al., 2007). In this study we compared the classification of putative *H. influenzae* strains using 1) the ME tree based on DNA sequences (McCrea et al., 2008), 2) DNA hybridization based 'library-on-a-slide' microarrays probed with conserved and variable regions of *iga*, 3) genomic dotblot hybridization with conserved and variable regions of *iga*, and 4) P6 7F3 epitope immunoblots.

The library-on-a-slide microarray (Zhang et al., 2004) consisted of the genomic DNA of 393 *Haemophilus* strains from three collections isolated from the throats of children in day care centers (Farjo et al., 2004, Granoff et al., 1979, St Sauver et al., 2000), 50 of which were also included in the ME tree. DNA was extracted by sonication (Zhang et al., 2005) and spotted on a microarray slide. For use as probes, the conserved β-core region of *iga* (Kilian, 2005) was amplified from *H. influenzae* strain 26225 using primers (*iga*BF1 5'-TGAATAACGAGGGGCAATATAAC-3' and *iga*BR2 5'-TCACCGCACTTAATCACTGAAT-3') corresponding to bases 4124–4978 in strain HK368 (GenBank accession M87492). Also, the variable region of *iga* was amplified from *H. influenzae* type b strain Eagan using primers (*iga*LF1 5'-GTTCCACCACCTGCGCCTGCTAC-3' and *iga*LR2 5'-GTTATATTGCCCCTCGTTATTCAT-3') corresponding to bases 3334–4146 in HK368 (Vitovski et al., 2002). The variable and conserved region fragments were fluorescein labeled as described (Kong et al., 2006). Duplicate microarrays were spotted, probed, and washed as described (Kong et al., 2006), with these exceptions: hybridization and washes were at 65 °C, PerfectHyb Plus (Sigma-Aldrich, St. Louis, MO) was used as a prehybridization and hybridization buffer, and the DNA concentration-control probe was a mixture of seven *H. influenzae* MLST gene fragments (http://Haemophilus.mlst.net/) and the coding region of *pepN*, in equal concentrations. Arrays hybridized with the conserved *iga* region, used a digoxigenin (DIG) labeled (DIG High Prime, Roche, Indianapolis) concentration-control probe, and then slides were destained and stripped by washing successively with 100% ethanol, 4M NaOH, 2 X SSC twice, and rehybridized with a fluorescein-labeled conserved region *iga* probe. Arrays hybridized with the variable *iga* region used a fluorescein-labeled concentration-control probe or a fluorescein-labeled *iga* variable region probe.

The Spotfinder v.3.1.1 program (http://www.tm4.org/spotfinder.html) determined the intensity of hybridization to each spot (settings: histogram, mask size 10, spot size 16, top background cut off 5%). MIDAS v.2.19 (http://www.tm4.org/midas.html) was used to remove poorly hybridized spots and to control for the concentration of genomic DNA in each spot by comparing the *iga* probing intensity with the concentration-control intensity.

Programs written in the statistical software "R" (http://www.r-project.org/) graphed the frequency of the log transformed *iga* to concentration-control signal ratio ($\log_2$ (I(*iga*)/I

(control)) where I = intensity calculated by Spotfinder) (Figure 1). Given that some strains contain the gene and some do not, thus the bimodal nature of the graphs, positive and negative spots clearly formed two different normal distributions. In cases in which two distributions were clearly separated (as in Figure 1), spots were considered positive for the probe if they were within the right-most peak (higher ratio) and negative if they were within the peak to the left (lower ratio). An obvious separation, however, between the positive and negative peaks for the two slides hybridized with the *iga* variable region probe was not seen. For these two slides, an alternate quantitative method was used (Figure 2). A two-component Gaussian mixture model was fitted using the EM algorithm (Dempster et al., 1977) to classify the observed intensities into positive or negative spots. This procedure was performed using an R program and is identical to the approach developed by Fraley and Raftery (2003) implemented in an R software package MCLUST (http://www.stat.washington.edu/mclust). A >50% probability was used as the cut off between positive and negative hybridization results. Although the absolute intensities of the spots varied, each strain was consistently positive or negative on duplicate slides.

Variable region *iga* dotblot analysis and P6 immunoblots were previously published (McCrea et al., 2008), for 197 strains that were also included in the ME tree. Conserved region *iga* dotblot analyses were performed as described (Xie et al., 2006) for 118 strains also in the ME tree. The conserved region *iga* β-core domain-specific probe was amplified from the *H. influenzae* strain Eagan using primers described above.

The ability of the three nonphylogenetic techniques to correctly identify the strains, defined as *H. haemolyticus* and *H. influenzae* based on the ME tree, is shown on Table 1. The *iga* variable region probe used in the dotblot hybridization assay, amplified from the *H. influenzae* strain Rd, was the best approximation to the ME tree, correctly categorizing 100% of the strains analyzed by both methods (McCrea et al., 2008). The microarray hybridization technique using a similar *iga* variable region probe, amplified from the *H. influenzae* strain Eagan, correctly categorized 96% of the strains. Using the microarray technique, 331 (84%) of the 393 total strains, including strains not used in the ME tree, were positive for the conserved region *iga* probe. When probed with the variable region of *iga*, 298 (76%) were positive (Table 2). Cross-species conservation of the conserved region of *iga* may account for the greater number of *H. haemolyticus* strains that hybridized to it.

Of rapid methods tested, dotblots probed with the *iga* variable region correlated best with the results of the multigene phylogenetic analysis (McCrea et al., 2008). However, the same region of *iga*, used as a microarray probe, was nearly as accurate, lower in cost, and less time-consuming per strain because more strains could be analyzed at once. The cost of consumable reagents to analyze one isolate using the phylogenetic analysis is about 30% less than the cost for one microarray slide that contains hundreds of isolates (data not shown). The hands-on time required to examine hundreds of strains on an automated microarray is only about three times that of running one sample through the phylogenetic analysis (unpublished observation), and the number of trees to be analyzed increases rapidly with the number of strains (Felsenstein, 1978). Because nasopharyngeal colonization is frequently polyclonal (Farjo et al., 2004, Mukundan et al., 2007, St Sauver et al., 2000), our laboratory now analyzes up to 30 *H. influenzae*-like colonies per throat swab to attempt to sample most of the strains. Even small studies quickly accumulate many isolates. For studies analyzing a large number of isolates, high throughput methods such as microarrays are time efficient and cost effective.

The choice between phylogenetic analysis and microarray analysis to identify a strain as *H. influenzae* or *H. haemolyticus* depends, in part, on the questions asked in designing the study at hand. A representation of the history of the relationships between strains is a by-product of the phylogenetic method, which would be useful for future genetic comparisons and

evolutionary studies. Additional identical microarray slides may be printed for little additional cost, which would be useful for future studies of variation in genetic regions between strains. If only a few strains need to be analyzed, we would recommend retrieving the sequences used by McCrea et. al. (2008) from Genbank, sequencing the five genes in the new strains, and building a phylogenetic tree. The cost/time/future-usefulness decision will be different for each research laboratory.

## Acknowledgements

Funding was provided by grants from the National Institute of Deafness and Other Communication Disorders (DC 05840) and the National Heart, Lung, and Blood Institute (HL 083893) awarded to J.R.G.

## References

Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the *EM* algorithm. J. Roy. Stat. Soc. B 1977;39:1–38.

Farjo RS, Foxman B, Patel MJ, Zhang L, Pettigrew MM, McCoy SI, Marrs CF, Gilsdorf JR. Diversity and sharing of *Haemophilus influenzae* strains colonizing healthy children attending day-care centers. Pediatr. Infect. Dis. J 2004;23:41–46. [PubMed: 14743045]

Felsenstein J. The number of evolutionary trees. Syst. Zool 1978;27:27–33.

Fraley C, Raftery AE. Enhanced model-based clustering, density estimation, and discriminant analysis software. MCLUST. J. Classif 2003;20:263–286.

Granoff DM, Gilsdorf J, Gessert C, Basden M. *Haemophilus influenzae* type B disease in a day care center: eradication of carrier state by rifampin. Pediatrics 1979;63:397–401. [PubMed: 312483]

Juliao PC, Marrs CF, Xie JP, Gilsdorf JR. Histidine auxotrophy in commensal and disease-causing nontypeable *Haemophilus influenzae*. J. Bacteriol 2007;189:4994–5001. [PubMed: 17496076]

Kilian, M. Genus III. *Haemophilus* Winslow, Broadhurst, Buchanan, Krumwiede, Rogers and Smith1917, 561$^{AL}$. In: Garrity, GM., editor. Bergey's Manual of Systematic Bacteriology. New York, NY: Springer-Verlag; 2005. p. 883-904.

Kilian M. A taxonomic study of the genus *Haemophilus*, with the proposal of a new species. J. Gen. Microbiol 1976;93:9–62. [PubMed: 772168]

Kong Y, Cave MD, Zhang L, Foxman B, Marrs CF, Bates JH, Yang ZH. Population-based study of deletions in five different genomic regions of *Mycobacterium tuberculosis* and possible clinical relevance of the deletions. J. Clin. Microbiol 2006;44:3940–3946. [PubMed: 16957037]

McCrea KW, Xie J, LaCross N, Patel M, Mukundan D, Murphy TF, Marrs CF, Gilsdorf JR. Relationships of nontypeable *Haemophilus influenzae* strains to hemolytic and nonhemolytic *Haemophilus haemolyticus* strains. J. Clin. Microbiol 2008:406–416. [PubMed: 18039799]

Mukundan D, Ecevit Z, Patel M, Marrs CF, Gilsdorf JR. Pharyngeal colonization dynamics of *Haemophilus influenzae* and *Haemophilus haemolyticus* in healthy adult carriers. J. Clin. Microbiol 2007;45:3207–3217. [PubMed: 17687018]

Murphy TF, Brauer AL, Sethi S, Kilian M, Cai X, Lesse AJ. *Haemophilus haemolyticus*: a human respiratory tract commensal to be distinguished from *Haemophilus influenzae*. J. Infect. Dis 2007;195:81–89. [PubMed: 17152011]

St Sauver J, Marrs CF, Foxman B, Somsel P, Madera R, Gilsdorf JR. Risk factors for otitis media and carriage of multiple strains of *Haemophilus influenzae* and *Streptococcus pneumoniae*. Emerg. Infect. Dis 2000;6:622–630. [PubMed: 11076721]

Vitovski S, Dunkin KT, Howard AJ, Sayers JR. Nontypeable *Haemophilus influenzae* in carriage and disease: a difference in IgA1 protease activity levels. JAMA 2002;287:1699–1705. [PubMed: 11926894]

Xie J, Juliao PC, Gilsdorf JR, Ghosh D, Patel M, Marrs CF. Identification of new genetic regions more prevalent in nontypeable *Haemophilus influenzae* otitis media strains than in throat strains. J. Clin. Microbiol 2006;44:4316–4325. [PubMed: 17005745]

Zhang L, Foxman B, Gilsdorf JR, Marrs CF. Bacterial genomic DNA isolation using sonication for microarray analysis. BioTechniques 2005;39:640–644. [PubMed: 16315370]

Zhang L, Srinivasan U, Marrs CF, Ghosh D, Gilsdorf JR, Foxman B. Library on a slide for bacterial comparative genomics. BMC Microbiol 2004:4–12. [PubMed: 14733619]

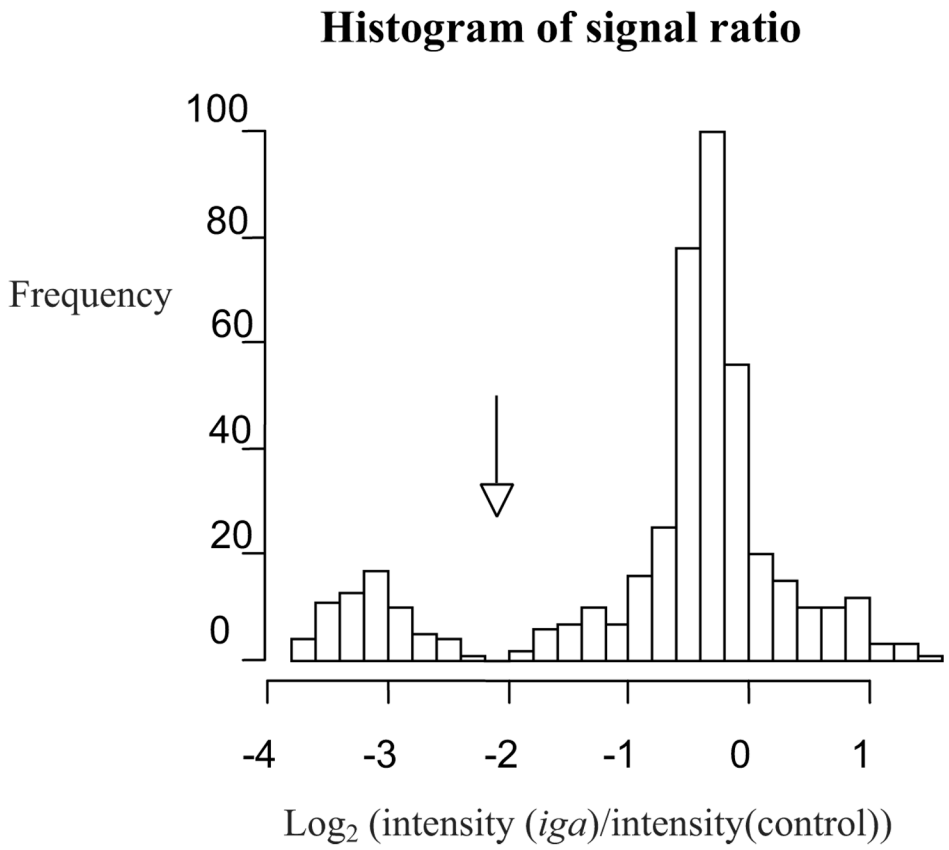


**Figure 1.**
Histogram used to determine cut off between positive and negative strains for *iga*-conserved probe on microarray. Strains positive for the probe are to the right of arrow, negative to the left. Intensities calculated using Spotfinder.

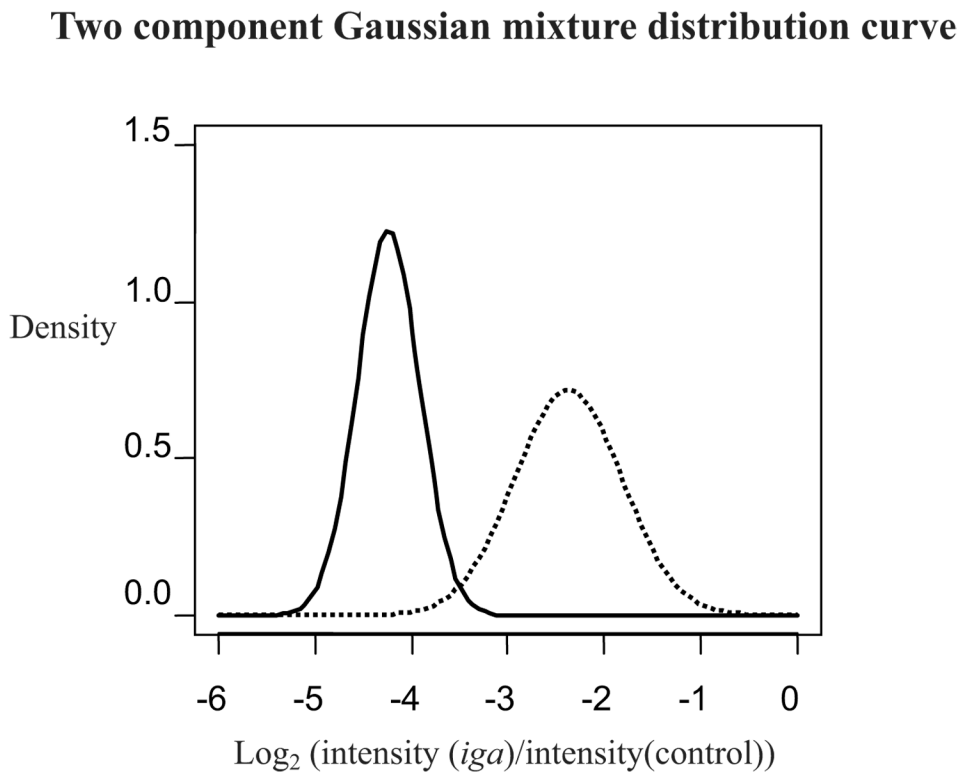


**Figure 2.**
Modeled intensities were used when histogram was unclear. The solid line indicates modeled curve for strains negative for the *iga*-variable probe, the dotted line for positive strains. A signal at the intersection of the two lines has a 50% probability of being positive. Strains with signals to the right of the intersection were considered positive, to the left negative.

## Table 1

Methods to distinguish *H. influenzae* from *H. haemolyticus*, as defined by five-gene ME tree clusters[a]

| | A: Microarray hybridization using *Iga* probes<br>Hh[b] cluster<br>N = 25 strains | <br>Hi[c] cluster<br>N = 25 strains |
|---|---|---|
| Variable *iga* +[d]<br>Variable *iga* –[f] | 0<br>**25** | **23**[e]<br>2 |
| Conserved *iga* +<br>Conserved *iga* − | 8<br>**17** | **23**<br>2 |

| | B: Dotblot hybridization using *Iga* probes<br>Hh cluster | <br>Hi cluster |
|---|---|---|
| Variable *iga* +[*]<br>Variable *iga* −[*] | 0 (N = 109 strains)<br>**109** | **88** (N = 88 strains)<br>0 |
| Conserved *iga* +<br>Conserved *iga* − | 11 (N = 42 strains)<br>**31** | **74** (N = 76 strains)<br>2 |

| | C: Immunoblot using 7F3 antibody of P6[*]<br>Hh cluster<br>N = 109 strains | <br>Hi cluster<br>N = 88 strains |
|---|---|---|
| P6 +<br>P6 − | 13<br>**96** | **85**<br>3 |

[a] Comparisons are to clusters of strains in the "gold standard " tree (McCrea et al., 2008).

[b] Hh designates *Haemophilus haemolyticus*.

[c] Hi designates *H. influenzae*.

[d] + indicates positive hybridization.

[e] Bold numbers indicate correctly categorized strains.

[f] – indicates no hybridization.

[*] from published data (McCrea et al., 2008)

**Table 2**

Statistics comparing five-gene tree results to other methods

| | % Strains Misassigned[a] | % + Predictive Value[b] | % − Predictive Value[c] | % Sensitivity[d] | % Specificity[e] |
|---|---|---|---|---|---|
| P6 7F3 antibody* | 8.1 | 86.7 | 97 | 96.6 | 88.1 |
| *Iga* conserved dotblot | 11 | 87 | 94 | 97 | 74 |
| *Iga* variable* dotblot | 0 | 100 | 100 | 100 | 100 |
| *Iga* variable microarray | 4 | 100 | 92.6 | 92 | 100 |
| *Iga* conserved microarray | 20 | 74.2 | 89.5 | 92 | 68 |

[a] Misassigned strains: *H. haemolyticus* positive for *iga* or p6 and *H. influenzae* negative for *iga* or P6.

[b] Positive predictive value: correctly identified positives divided by all strains that were positive for that probe.

[c] Negative predictive value: correctly identified negatives divided by all strains that were negative for that probe.

[d] Sensitivity: positive Hi strains divided by all Hi strains.

[e] Specificity: negative Hh strains divided by all Hh strains.

* from published data, (McCrea et al., 2008).