AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Clinical Microbiology

Comparative Genomic Analysis of *Haemophilus haemolyticus* and Nontypeable *Haemophilus influenzae* and a New Testing Scheme for Their Discrimination

Fang Hu,[a] Lavanya Rishishwar,[b,c,d] Ambily Sivadas,[b] Gabriel J. Mitchell,[b,e] I. King Jordan,[b,c,d] Timothy F. Murphy,[f] Janet R. Gilsdorf,[g] Leonard W. Mayer,[a] Xin Wang[a]

Meningitis and Vaccine Preventable Diseases Branch, Division of Bacterial Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA[a]; School of Biology, Georgia Institute of Technology, Atlanta, Georgia, USA[b]; Applied Bioinformatics Laboratory, Atlanta, Georgia, USA[c]; PanAmerican Bioinformatics Institute, Cali, Valle del Cauca, Colombia[d]; Institute of Science and Technology Austria, Klosterneuburg, Austria[e]; Division of Infectious Diseases, Department of Medicine, Clinical and Translational Research Center, Department of Microbiology and Immunology, University at Buffalo, State University of New York, Buffalo, New York, USA[f]; Department of Pediatrics, University of Michigan, Ann Arbor, Michigan, USA[g]

***Haemophilus haemolyticus* has been recently discovered to have the potential to cause invasive disease. It is closely related to nontypeable *Haemophilus influenzae* (NT *H. influenzae*). NT *H. influenzae* and *H. haemolyticus* are often misidentified because none of the existing tests targeting the known phenotypes of *H. haemolyticus* are able to specifically identify *H. haemolyticus*. Through comparative genomic analysis of *H. haemolyticus* and NT *H. influenzae*, we identified genes unique to *H. haemolyticus* that can be used as targets for the identification of *H. haemolyticus*. A real-time PCR targeting *purT* (encoding phosphoribosylglycinamide formyltransferase 2 in the purine synthesis pathway) was developed and evaluated. The lower limit of detection was 40 genomes/PCR; the sensitivity and specificity in detecting *H. haemolyticus* were 98.9% and 97%, respectively. To improve the discrimination of *H. haemolyticus* and NT *H. influenzae*, a testing scheme combining two targets (*H. haemolyticus purT* and *H. influenzae hpd*, encoding protein D lipoprotein) was also evaluated and showed 96.7% sensitivity and 98.2% specificity for the identification of *H. haemolyticus* and 92.8% sensitivity and 100% specificity for the identification of *H. influenzae*, respectively. The dual-target testing scheme can be used for the diagnosis and surveillance of infection and disease caused by *H. haemolyticus* and NT *H. influenzae*.**

Haemophilus haemolyticus is a human commensal that colonizes the respiratory tract. It shares high similarities in morphology, biochemistry, and genetics with nontypeable *Haemophilus influenzae* (NT *H. influenzae*) (1–3). NT *H. influenzae* has emerged as the major cause of invasive *H. influenzae* disease since the implementation of *H. influenzae* serotype b (*H. influenzae* b) vaccine (4). NT *H. influenzae* infections such as childhood otitis media and respiratory tract infections in adults with chronic obstructive pulmonary disease (COPD) result in an enormous social and economic burden to societies (5). *H. haemolyticus* was recently reported to cause invasive disease (1). These cases were previously misidentified as NT *H. influenzae* due to the lack of proper identification methods to discriminate between the two species (6). Reevaluation of NT *H. influenzae* cases from previous years revealed that about 2% (7/374) of the NT *H. influenzae* invasive cases reported from Active Bacterial Core surveillance (ABCs) were in fact caused by *H. haemolyticus* (1).

Misidentification of *H. haemolyticus* as NT *H. influenzae* has been repeatedly reported by multiple research groups, with up to 40% of isolates being misidentified as NT *H. influenzae* using classical phenotypic methods in clinical microbiology laboratories (1, 7–11). The misidentification of *H. haemolyticus* has a potential impact on accurate assessment of the prevalence of antibiotic-resistant NT *H. influenzae* (12). Therefore, a rapid and accurate method is needed to distinguish *H. haemolyticus* from NT *H. influenzae* and to better understand the epidemiology of *H. haemolyticus* infections. Testing schemes, including both standard microbiological and molecular methods, have been proposed to improve the identification of the two bacteria. These molecular methods target a number of genes such as the lipo-oligosaccharide gene *lgtC*, the IgA protease gene *iga*, heme acquisition genes (*hxuABC*, *hemR*, and *hup*), the fuculose kinase gene *fucK*, and the *Haemophilus* protein D gene (*hpd*), with *hpd*-based PCR being the best method for discriminating NT *H. influenzae* and *H. haemolyticus* (13–16). However, identification of *H. haemolyticus* relies on negative results, as all these genes are present in *H. influenzae* strains and absent in the majority of *H. haemolyticus* strains. A test for specific identification of *H. haemolyticus* is needed to better understand the epidemiology of *H. haemolyticus* disease. In this study, we conducted comparative genomic analysis to identify unique target genes that are exclusively present in *H. haemolyticus* strains and can be used to develop PCR assays for specific detection and identification of *H. haemolyticus*.

Received 13 July 2016 **Returned for modification** 3 August 2016
Accepted 26 September 2016

Accepted manuscript posted online 5 October 2016

Citation Hu F, Rishishwar L, Sivadas A, Mitchell GJ, Jordan IK, Murphy TF, Gilsdorf JR, Mayer LW, Wang X. 2016. Comparative genomic analysis of *Haemophilus haemolyticus* and nontypeable *Haemophilus influenzae* and a new testing scheme for their discrimination. J Clin Microbiol 54:3010–3017. doi:10.1128/JCM.01511-16.

Editor: N. A. Ledeboer, Medical College of Wisconsin

Address correspondence to Xin Wang, gqe8@cdc.gov.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/JCM.01511-16.

TABLE 1 *H. haemolyticus* and *H. influenzae* strains, phenotypes, and whole-genome sequences analyzed in this study

Species and strain	NCBI accession no.	Path[a]	Hemolysis[b]	No. of contigs[c]	Total length (bases)[d]	Largest contig (bases)[e]	N_{50} (bases)[f]
H. haemolyticus							
M19107	AFQN00000000.1	No	Yes	115	1,909,361	81,211	30,635
M19501	AFQO00000000.1	No	No	22	1,884,911	364,637	238,571
M21127	AFQP00000000.1	Yes	Yes	34	1,943,222	319,487	121,923
M21621	AFQQ00000000.1	Yes	Yes	28	2,089,205	339,425	234,852
M21639	AFQR00000000.1	Yes	No	50	2,326,092	187,351	98,989
H. influenzae							
PittHH	AAZH00000000.1	Yes	No	51	1,831,290	164,041	95,398
R3021	AAZE00000000.1	No	No	45	1,875,022	314,724	79,720
22.4-21	AAZJ00000000.1	No	No	41	1,848,720	365,947	129,943
PittAA	AAZG00000000.1	Yes	No	38	1,875,062	245,610	123,700
M21709		Yes	No	34	1,822,150	319,861	141,272
NT127	ACSL00000000.1	Yes	No	34	1,860,733	210,220	77,084
6P18H1	ABWW00000000.1	Yes	No	28	1,912,236	392,297	152,038
PittII	AAZI00000000.1	Yes	No	22	1,952,226	385,065	135,736
3655	AAZF00000000.1	Yes	No	21	1,876,504	381,682	184,545
7P49H1	ABWV00000000.1	Yes	No	18	1,827,137	550,114	368,081
22.1-21	AAZD00000000.1	No	No	15	1,886,604	648,211	440,092
10810	NC_016809	Yes	No	1	1,981,535	1,981,535	1,981,535
86-028NP	NC_007146	Yes	No	1	1,914,490	1,914,490	1,914,490
F3031	NC_014920	Yes	No	1	1,985,832	1,985,832	1,985,832
F3047	NC_014922	Yes	No	1	2,007,018	2,007,018	2,007,018
PittEE	NC_009566	Yes	No	1	1,813,033	1,813,033	1,813,033
PittGG	CP000672	Yes	No	1	1,887,192	1,887,192	1,887,192
R2846	NC_017452	Yes	No	1	1,819,370	1,819,370	1,819,370
R2866	NC_017451	Yes	No	1	1,932,306	1,932,306	1,932,306
Rd-KW20	NC_000907	Yes	No	1	1,830,138	1,830,138	1,830,138

[a] Whether or not strain was determined to be pathogenic (disease causing).
[b] Whether or not a strain was shown to be capable of hemolysis.
[c] Total number of contigs of ≥1 kbp (i.e., large contigs) produced in the assembly.
[d] Total number of bases assembled into large contigs.
[e] Size of the largest contig in the assembly.
[f] Size of the contig at which contigs of that length or longer cover at least half of the assembly.

MATERIALS AND METHODS

Bacterial strains and growth conditions. Bacterial strains used in this study included ATCC strains, clinical invasive isolates, and carriage isolates. Eighty-nine clinical invasive *H. influenzae* isolates were collected as part of the Active Bacterial Core surveillance (ABCs) of the Centers for Disease Control and Prevention's Emerging Infectious Program (http://www.cdc.gov/abcs/reports-findings/surv-reports.html). Seventy-eight *H. influenzae* isolates and 45 *H. haemolyticus* isolates were collected during a carriage survey in Minnesota in 2009 (17). Additional *H. haemolyticus* isolates were kindly provided by University at Buffalo, State University of New York ($n = 31$), and University of Michigan Medical Center ($n = 104$). The clinical and carriage isolates in this study have been characterized and confirmed using standard microbiology and 16S rRNA gene sequencing.

Whole-genome sequencing and comparison. Genome sequences of the five *H. haemolyticus* strains were used for comparison with those of *H. influenzae*. Details of whole-genome sequencing of the strains and the assembled sequences have been reported previously (18). Twenty complete and draft *H. influenzae* (19 NT *H. influenzae* and 1 *H. influenzae* b genome) genome sequences were downloaded from the NCBI RefSeq and Shotgun Assembly Sequence database (19) (Table 1). The evolutionary relationships among these *H. haemolyticus* and *H. influenzae* strains were characterized using analysis of 16S rRNA gene sequences and multilocus sequence typing (MLST) locus sequences. 16S rRNA and MLST locus sequences were obtained from the annotated GenBank entries at NCBI. Multiple copies of 16S rRNA loci were retained with postfix .a, .b, etc. 16S rRNA and concatenated sequences of the seven MLST loci were aligned using the program MUSCLE (20), and the resulting alignments were used to reconstruct phylogenetic trees using the neighbor-joining method (21) as implemented in the program MEGA (22). Evolutionary relationships among the *H. haemolyticus* and *H. influenzae* strains were also characterized using whole-genome sequences by computing the average nucleotide identity (ANI) (23, 24) between each pair of genomes using the program MUMmer (25).

Identification of *H. haemolyticus*-specific genes by comparative genomics. Complete gene (protein-coding) sets from the *H. haemolyticus* and *H. influenzae* genome sequences were obtained from the annotated GenBank entries at NCBI (Table 1). All-against-all nucleotide sequence comparison of complete gene sets was performed using the BLASTCLUST algorithm (26) using default parameters. BLASTCLUST creates clusters of genes that share sequence similarity and coverage above a given threshold using single-linkage clustering. The resulting clusters are exclusive, with a gene mapping to exactly one cluster. The results of the clustering procedure are visualized as a presence/absence matrix.

Sanger sequencing of target genes. Three genes (*purT*, coding for putative phosphoribosylglycinamide formyltransferase 2; *hdg*, coding for putative hydrogenase-2 small chain; and *sod*, coding for putative superoxide dismutase [Cu-Zn]-like) from five additional *Haemophilus haemolyticus* isolates were sequenced using the Sanger sequencing method to validate their sequence conservation levels. Primers for amplification and sequencing of target genes are shown in Table 2. The PCR amplification and DNA Sanger sequencing were performed as described previously us-

TABLE 2 Primers and probes used for DNA sequencing and real-time PCR assays[a]

Purpose and target gene	Primer designation	5′-3′ nucleotide sequence[d]	Amplicon length (bp)	Final concn (nM)
DNA amplification and sequencing primers				
purT[b]	*purT* Fw1	CACTTGGCACAGCATTAACC	1,145	400
	purT Fw	TAACCCCTAATGCAACCAAAGT	1,130	400
	purT Rv	CTTGTTTTACTTTTTCCAACGC		400
	purT seq	CCAATAGGTCATCGTCAAGAAGA		160
	purT seq1R	TCTTCTTGACGATGACCTATTGG		160
hdg	*hdg* Fw	TTATGAAATTATGTACCGCACTTG	1,084	400
	hdg Fw1	TGGCGTTTTTTCTGCRTTG	1,125	400
	hdg Rv	TCCTTGATCTTGATTGGCTTG		400
	hdg seq	AACMGTTGCCTATATCATCACTT		160
	hdg seq1R	AAGTGATGATATAGGCAACKGTT		160
sod	*sod* Fw	GTTAGTGCGGTATGTTCAGTTGGT	502	400
	sod Rv	CCAAGTGGMGCTGGATGATC		400
	sod seq Fw	GTTAGTGCGGTATGTTCAGTTGGT		160
	sod seq Rv	CCAAGTGGMGCTGGATGATC		160
RT-PCR primers and probes				
purT 1	*purT* Fw1	ATTAACCCCTAATGCAACCAAAGT	91	100
	purT Rv1	CTTCAACGCCTAAACGTTGCA		900
	purT pb1[c]	ATGTTGGGTTCTGGTGAA		300
purT 2	*purT* Fw2	TTAATGTTGTGCCGACTGCAA	78	100
	purT Rv2	AACTCTTCTGAAGCTAGACGACGAAT		600
	purT pb2[c]	TCAACTTACGATGAATCGTGA		300

[a] All primers and probes were synthesized at the CDC Biotechnology Core Facility.
[b] Primers for amplification and sequencing of *purT*.
[c] All probes were labeled with 6-carboxyfluorescein at the 5′ end and black hole quencher at the 3′ end.
[d] K = G+T; M = A+C; R = A+G.

ing chromosomal DNA template extracted from five additional isolates (27). The gene sequences were assembled using DNAStar Lasergene 9 (DNAStar, Inc., Madison, WI).

RT-PCR. Chromosomal DNA and crude cell lysates were prepared as described previously (28). Primer Express 3.0 (Applied Biosystems) was used to design appropriate primers and probes for detecting *purT*. All primers and probes used in this study were optimized by testing in the range of 100 to 900 nM and 100 to 300 nM, respectively. Real-time PCR (RT-PCR) was performed as described previously (28). A chromosomal DNA or crude cell lysate was considered positive if the cycle threshold (C_T) value was equal to or less than 35 and negative if the C_T value was greater than 40. If a C_T value was between 35 and 40, the sample was diluted 10-fold and retested to determine if PCR inhibitors were present. The specimen was considered positive if the C_T value of the diluted specimen was equal to or less than 35 and negative if the C_T value was greater than 35. To determine the lower limit of detection (LLD), genomic DNA was extracted from *H. haemolyticus* isolates (28). The DNA concentration was determined using a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE), adjusted to 20 ng/μl, and then 10-fold serially diluted in PCR-grade water. Each dilution was tested by PCR in triplicate. DNA concentration was converted to genome equivalents per microliter on the basis of 1.8 Mb (carriage isolate) and 2.0 Mb (invasive isolate) per *H. haemolyticus* genome (AFQN00000000, AFQO00000000, AFQP00000000, AFQQ00000000, AFQR00000000). The LLD for an RT-PCR assay was defined as the DNA concentration that yielded a C_T value of 35. Confidence intervals for the *H. haemolyticus* and *H. influenzae* sensitivity and specificity values, calculated based on the results of the RT-PCR assays, were determined using standard validation analyses. The validation analyses were performed using SAS 9.2 (SAS Institute Inc., Cary, NC), and exact binomial 95% confidence intervals were estimated using Stata v. 9.2 (StataCorp).

RESULTS

Comparative genomic analysis of *H. haemolyticus* and *H. influenzae*. Comparative analyses of 25 whole-genome sequences of *H. haemolyticus* (5) and *H. influenzae* (20) were performed in an effort to assess whether strains from these two species, which are often confused using classical phenotypic or biochemical methods, can be clearly distinguished at the genomic level. Phylogenetic analysis using individual (16S rRNA) or multiple (MLST) gene loci clearly distinguishes *H. haemolyticus* and *H. influenzae* evolutionary lineages (Fig. 1A and B). Comparison of whole-genome sequences using the ANI technique also yields unambiguous discrimination between the *H. haemolyticus* and *H. influenzae* strains (Fig. 1C). These results suggested the possibility that there may be individual gene loci with presence/absence patterns that distinguish *H. haemolyticus* from *H. influenzae*. Such loci would represent ideal targets for a real-time PCR (RT-PCR) typing scheme.

Identification of *H. haemolyticus*-specific RT-PCR gene targets. All-against-all comparison of complete gene sets from the 25 genomes analyzed here was used to define clusters of homologous genes that are exclusive to the *H. haemolyticus* lineage. A total of 93 clusters were found with homologous genes that were present in all 5 of the *H. haemolyticus* genomes and absent in all 20 of the *H. influenzae* genomes (Fig. 2). All of these clusters represent potential gene targets for *H. haemolyticus*-specific RT-PCR, and clusters with conserved flanking regions (for primer and probe binding sites) were considered for the development of RT-PCR assays. Three genes (*purT*, coding for putative phosphoribosylglycinamide formyltransferase 2; *hdg*, coding for putative hydroge-

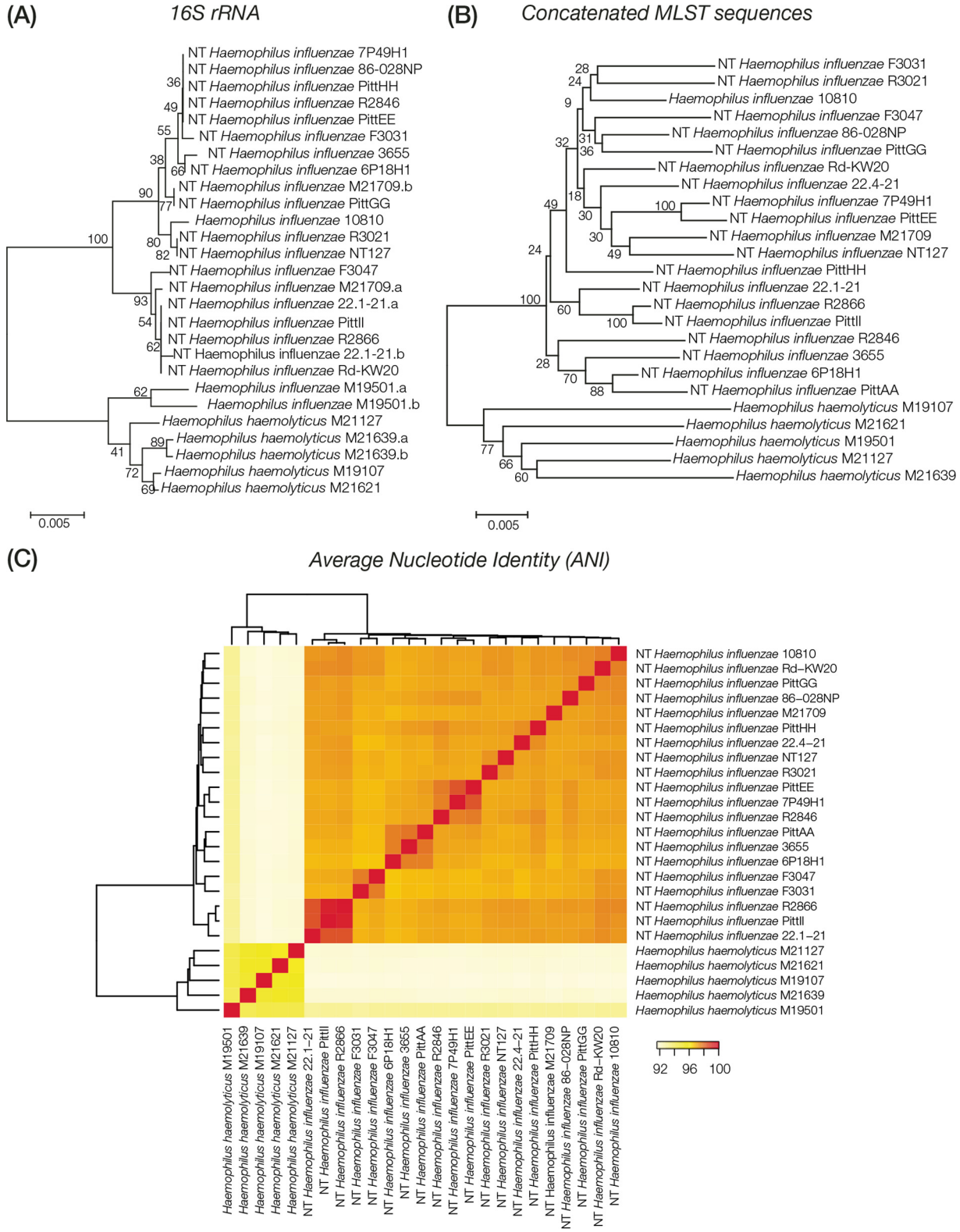

FIG 1 Comparison of *H. haemolyticus* versus *H. influenzae* evolutionary lineages. *H. haemolyticus* and *H. influenzae* strains analyzed here are labeled with their species abbreviations and the strain names shown in Table 1. (A and B) Phylogenetic trees based on 16S rRNA gene (A) and concatenated MLST loci (B) showing the evolutionary relationships of the *H. haemolyticus* and *H. influenzae* genome sequences analyzed here. Percent bootstrap values indicate support for internal nodes on the trees, and the branch length scale bars show *P* distances. (C) Results of whole-genome sequence comparisons among *H. haemolyticus* and *H. influenzae* based on ANI analysis. ANI values (percentages) between pairs of genomes are color coded as shown in the key, and the relationships among the genomes based on these values are shown as dendrograms on both axes.

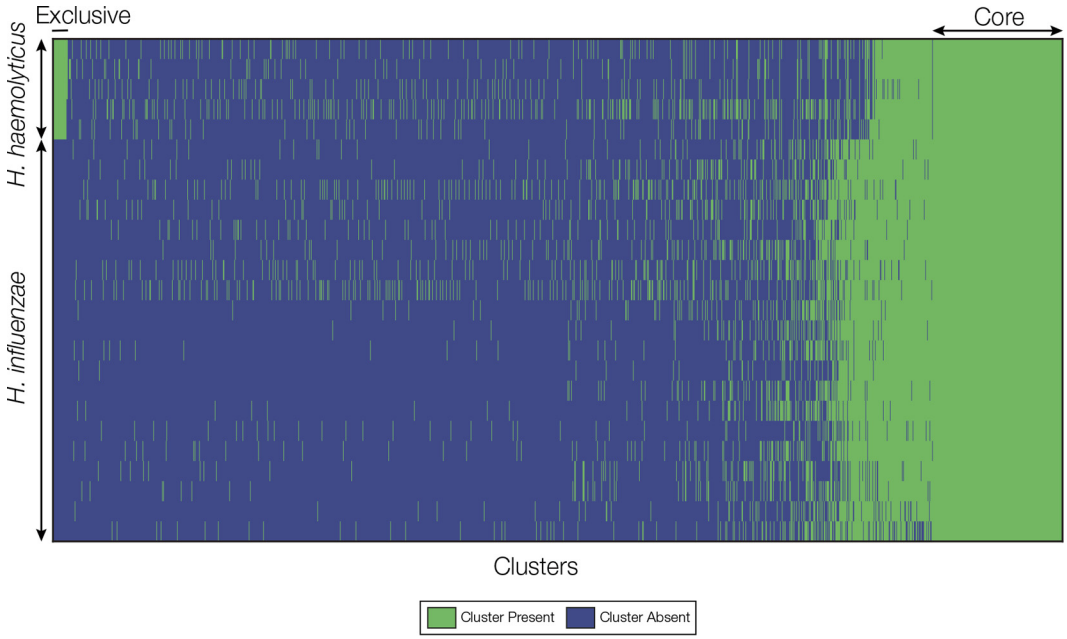

FIG 2 Homologous gene cluster presence/absence matrix. Genomes are shown as rows, and homologous gene clusters are shown as columns. The presence of a gene cluster in a genome is indicated by green, and cluster absence is indicated by blue. *H. haemolyticus* genomes are shown on top of the matrix, and *H. influenzae* genomes are shown below. Core clusters found in all genomes are shown on the right of the matrix, and clusters exclusive to *H. haemolyticus* (i.e., potential RT-PCR targets) are shown in the upper left corner of the matrix.

nase-2 small chain; and *sod*, coding for putative superoxide dismutase [Cu-Zn]-like) showed high sequence similarity among the 5 *H. haemolyticus* genomes (Fig. 3) and were further characterized using Sanger sequencing from 5 additional *H. haemolyticus* isolates in order to confirm the sequence similarity. Alignment of the 10 sequences for each of the 3 genes revealed a potential RT-PCR target *purT* (see Fig. S1 in the supplemental material); *purT* sequences showed more than 96% identity among the 10 *H. haemolyticus* strains. Two conserved regions within *purT* were selected for designing primers and probes for RT-PCR assays *purT* 1 and 2 (see Fig. S1 in the supplemental material). Sequences of the *purT* primers and probes were determined to be specific to *H. haemolyticus* by BLAST search against GenBank (Table 2).

Evaluation of the RT-PCR assays. Optimal concentrations of the RT-PCR primers and probes for the *purT* 1 and 2 assays are listed in Table 2. Of the 65 strains of non-*H. haemolyticus* bacterial species that were tested by the two *purT* assays (Table 3), none was positive for *purT*, suggesting that the assays were specific for *H. haemolyticus*. Both assays have consistently low average LLDs under the tested conditions, C_T values of 35 for 29 to 40 genomes/PCR, and an amplification curve with a higher plateau for assay 1. As a result, assay 1 was chosen for further validation using *H. haemolyticus* carriage and invasive strains.

A total of 347 strains (180 *H. haemolyticus* and 167 *H. influenzae* strains identified by 16S rRNA gene) were tested to determine the sensitivity and specificity of the *purT* 1 assay for the detection of *H. haemolyticus*. Of these strains, 178/180 *H. haemolyticus* strains (98.9% sensitivity) were positive for *purT*, 2 out of 180 *H. haemolyticus* strains were negative for *purT*, and 5 out of 167 *H. influenzae* strains were positive for *purT*. The *hpd* assay detects the *Haemophilus* protein D-encoding gene and is currently used for specific detection of *H. influenzae* regardless of the capsulation status (14). Using the same strain collection, the sensitivity and specificity of the *hpd* assay were 94% (157/167) and 97.8% (176/180) for the identification of *H. influenzae*, respectively.

A testing scheme combining the *purT* 1 and *hpd* assays was validated for distinguishing *H. haemolyticus* from *H. influenzae*, with $purT^+/hpd$ being the expected genotype for *H. haemolyticus* and $purT/hpd^+$ for *H. influenzae* (Fig. 4A). Of the 180 *H. haemolyticus* strains, 174 were positive for *purT* and negative for *hpd* ($purT^+/hpd$). Four were positive for *purT* and *hpd* ($purT^+/hpd^+$), and 2 were negative for both genes (*purT*/*hpd*), which are not the expected *H. haemolyticus* genotypes. Of the 167 *H. influenzae* strains, 155 were negative for *purT* and positive for *hpd* ($purT/hpd^+$). Two were $purT^+/hpd^+$, 7 were *purT*/*hpd*, and 3 were $purT^+/hpd$, which are not the expected genotypes for *H. influenzae*. Using 16S rRNA gene sequencing as the reference standard, the sensitivity and specificity of the testing scheme were 96.7% and 98.2% for the detection of *H. haemolyticus*, respectively, and 92.8% and 100% for the detection of *H. influenzae*, respectively (Fig. 4B).

DISCUSSION

Historically, *H. influenzae* has been the most important species of *Haemophilus* causing invasive human disease and fatalities (4). Bacteremia caused by other *Haemophilus* species has not been frequently reported. *H. haemolyticus* was recently reported to cause invasive disease in the United States (1). Little is known about the mechanisms of *H. haemolyticus* pathogenicity. *H. haemolyticus* may cause disease more as an opportunistic pathogen. By comparative genomic analyses, we did not find any genes coding for the biosynthesis of capsular polysaccharide in *H. haemolyticus* strains, a major virulence factor that confers bacterial resistance to phagocytosis and complement-mediated host defense.

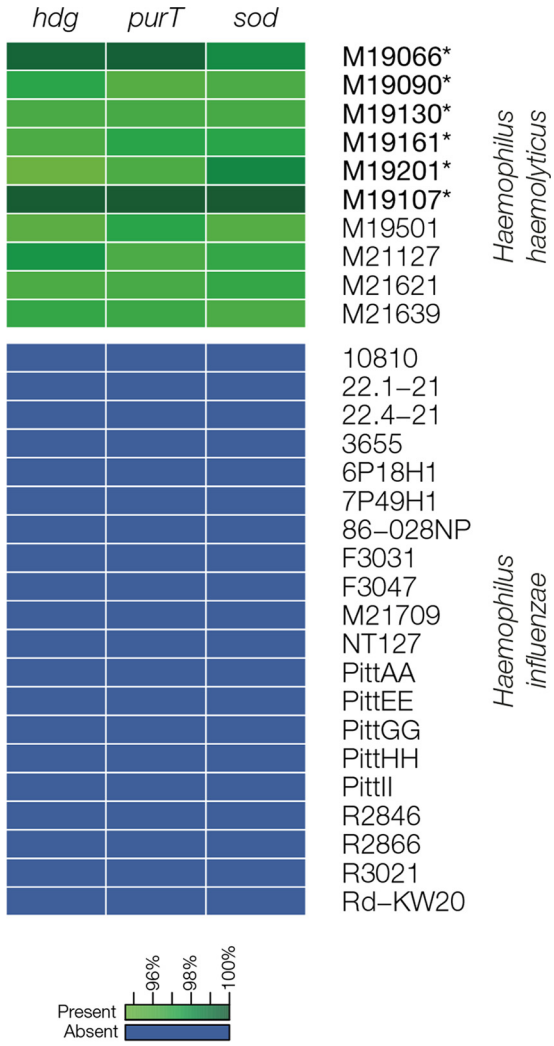

FIG 3 *H. haemolyticus*-specific potential RT-PCR target genes. The presence (green) and absence (blue) of 3 *H. haemolyticus*-specific genes—*hdg*, *purT*, and *sod*—are shown for the *H. haemolyticus* ($n = 10$) and *H. influenzae* ($n = 20$) strains analyzed here. Strain names are as shown in Table 1. The percent identities of *H. haemolyticus*-specific genes to reference genes from strain M19107 are color coded as shown in the key. The sequences of *hdg*, *purT*, and *sod* of six *H. haemolyticus* strains (marked with asterisks) were determined by Sanger sequencing. Five *H. haemolyticus* strains were sequenced by WGS. The M19107 strain was characterized by both methods.

However, genes encoding factors for colonization and invasion such as pili and IgA1 protease are present in these strains.

Unambiguous discrimination of the closely related species *H. influenzae* and *H. haemolyticus* is important for laboratory diagnosis and surveillance of *H. influenzae* and *H. haemolyticus* disease and carriage evaluations. Over the last decade, considerable research effort has focused on identifying molecular targets and suitable methodologies to differentiate NT *H. influenzae* from *H. haemolyticus*. One of the principal phenotypic differences between *H. haemolyticus* and NT *H. influenzae* is hemolysis by *H. haemolyticus*. However, this difference is often unreliable as *H. haemolyticus* can lose the defining hemolytic phenotype upon *in vitro* passage (1–3, 11, 29–31). A number of genetic targets (*hpd*, *ompP2*, *opmP6*, *lgtC*, 16S rRNA, *fucK*, and *iga*) have been evaluated for differentiation of NT *H. influenzae* from *H. haemolyticus* (15). However, no single gene target tested was able to unequivocally differentiate NT *H. influenzae* and *H. haemolyticus*. For example, 16S rRNA gene PCR permitted identification of *H. haemolyticus* and NT *H. influenzae* for only 90% of strains (6), while the *ompP6*-based assay detects about 97% of NT *H. influenzae* strains but also detects 12% of *H. haemolyticus* strains (2). Identification of *H. haemolyticus*-specific targets was challenging due to the lack of available *H. haemolyticus* genome sequences.

Whole-genome sequencing (WGS) provides a valuable tool for better understanding *H. haemolyticus* genetics and for identifying unique genetic targets for assay development. Recently, five *H. haemolyticus* strains were sequenced and annotated by our group (18). Comparative analysis revealed several *H. haemolyticus*-specific genes, including the *purT* gene; these genes were found to be present exclusively among 5 *H. haemolyticus* genomes and absent from 20 *H. influenzae* genomes. *purT* sequences are highly conserved among all of the *H. haemolyticus* strains examined in this study, which makes it an ideal target for RT-PCR development. While the *purT* assay was found to be highly sensitive and specific for *H. haemolyticus* detection, the addition of testing for the presence of *hpd* might provide enhanced discrimination for the detection of both organisms and could be used to enhance diagnosis and surveillance of *H. haemolyticus* and *H. influenzae* infections.

A few exceptions were observed (Fig. 4A): four *H. haemolyticus* and two *H. influenzae* strains were positive for both *purT* and *hpd* genes. Because of the high frequency of horizontal gene transfer

TABLE 3 Non-*H. haemolyticus* bacterial species tested in this study

Organism	No. of isolates
Branhamella catarrhalis	1
Corynebacterium diphtheriae	1
Cryptococcus neoformans	1
Haemophilus influenzae biogroup aegyptius	7
Haemophilus aphrophilus[a]	2
Haemophilus influenzae a, b, c, d, e, f, and NT	1 each
Haemophilus parahaemolyticus	1
Haemophilus parainfluenzae	6
Haemophilus paraphrophilus[a]	1
Haemophilus spp.	1
Neisseria meningitidis serogroups	
A, C, W, X, Y, Z, NG, and E	1 each
B and Z	2 each
Bordetella pertussis	1
Escherichia coli strain O16:K1	1
Escherichia coli strain O7:K1	1
Neisseria cinerea	2
Neisseria gonorrhoeae	2
Neisseria lactamica	12
Neisseria sicca	1
Neisseria subflava	1
Pseudomonas aeruginosa	1
Staphylococcus aureus	1
Streptococcus agalactiae	1
Streptococcus pneumoniae	1
Total	65

[a] Both *H. aphrophilus* and *H. paraphrophilus* have been classified as *Aggregatibacter aphrophilus* (37).

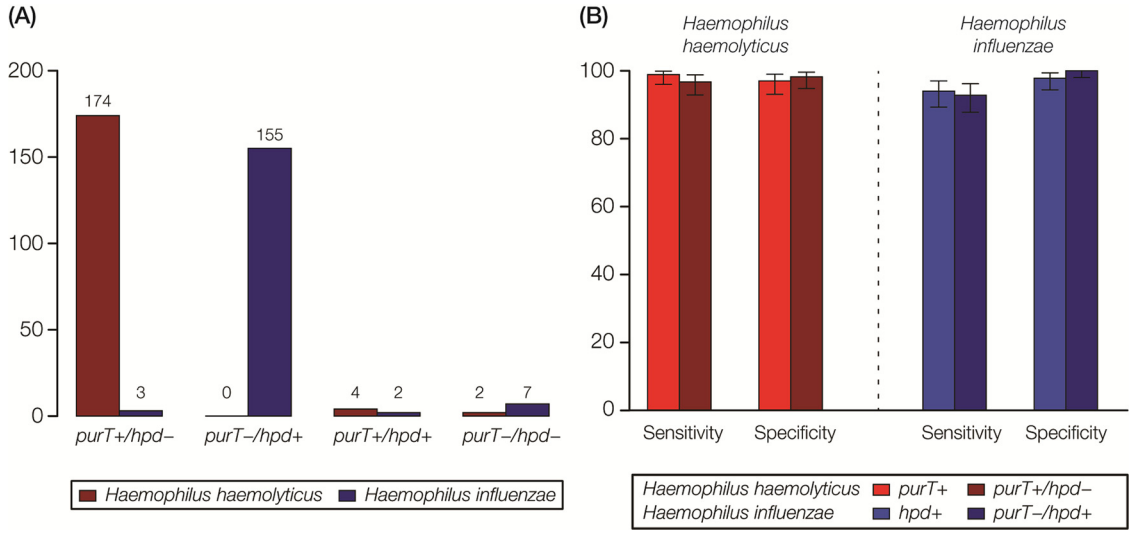

FIG 4 Results of RT-PCR assays for the discrimination of *H. haemolyticus* versus *H. influenzae* strains. (A) Four possible results for combined *purT* and *hpd* RT-PCR assays are evaluated: $purT^{+}/hpd$, $purT/hpd^{+}$, $purT^{+}/hpd^{+}$, and *purT*/*hpd*. The numbers of *H. haemolyticus* (red) and *H. influenzae* (blue) strains with each combination are shown. (B) Sensitivity and specificity values for discrimination of *H. haemolyticus* and *H. influenzae* based on *purT* or *hpd* RT-PCR assays alone compared to combined RT-PCR assays for *H. haemolyticus* ($purT^{+}/hpd$) and *H. influenzae* ($purT/hpd^{+}$).

among and between *Haemophilus* species, it is conceivable that, during pharyngeal colonization with both species, *H. haemolyticus* strains can acquire *H. influenzae* genes and vice versa. Two *H. haemolyticus* and seven *H. influenzae* strains were negative for both *purT* and *hpd* genes. The results seen with *purT* and *hpd* PCR-negative isolates may have been due to deletions at that locus or to sequence variation at any of the primer or probe binding sites (30, 32).

Multiple studies have shown that single-target-based tests are not ideal for discriminating bacterial species, particularly for closely related species, or for classifying the variant strains that have diverged from their species, such as fuzzy species of *Neisseria* (2, 15, 33). While multitarget-based approaches can improve the sensitivity and specificity, developing multiple tests for each bacterial species is a huge undertaking. WGS can potentially serve as a universal high-throughput method to improve the accuracy of species delineation and has been proven accurate for bacterial species classification (34). In addition, WGS provides much-enriched genetic information for strain subtyping and will be very useful for disease surveillance and outbreak investigations. Matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) is increasingly used for species classification in diagnostic microbiology laboratories. However, it may not be able to provide sufficient resolution for subtyping of bacterial pathogens, which is often performed in public health microbiology laboratories for disease surveillance and other large epidemiological surveys. The power of MALDI-TOF for identification of *H. haemolyticus* and *H. influenzae* is highly dependent on a well-defined reference spectrum for *H. haemolyticus* in species databases, which varies between laboratories (35, 36). Its utility in discriminating between *H. haemolyticus* and *H. influenzae* remains to be further validated.

As the laboratory bioinformatics capacity increases and costs of system acquisition decrease, the advanced molecular detection tools such as WGS and MALDI-TOF may be widely utilized in diagnostic and public health microbiology laboratories. However, rapid and less expensive methods with high throughput, such as PCR, remain valuable today in these laboratories for diagnosis and surveillance of infectious diseases.

ACKNOWLEDGMENTS

We are grateful to ABCs for providing strains and the Bacterial Meningitis Laboratory for technical support.

REFERENCES

1. **Anderson R, Wang X, Briere EC, Katz LS, Cohn AC, Clark TA, Messonnier NE, Mayer LW.** 2012. *Haemophilus haemolyticus* isolates causing clinical disease. J Clin Microbiol **50:**2462–2465. http://dx.doi.org/10.1128/JCM.06575-11.
2. **McCrea KW, Xie J, LaCross N, Patel M, Mukundan D, Murphy TF, Marrs CF, Gilsdorf JR.** 2008. Relationships of nontypeable *Haemophilus influenzae* strains to hemolytic and nonhemolytic *Haemophilus haemolyticus* strains. J Clin Microbiol **46:**406–416. http://dx.doi.org/10.1128/JCM.01832-07.
3. **Kilian M.** 1976. A taxonomic study of the genus Haemophilus, with the proposal of a new species. J Gen Microbiol **93:**9–62. http://dx.doi.org/10.1099/00221287-93-1-9.
4. **MacNeil JR, Cohn AC, Farley M, Mair R, Baumbach J, Bennett N, Gershman K, Harrison LH, Lynfield R, Petit S, Reingold A, Schaffner W, Thomas A, Coronado F, Zell ER, Mayer LW, Clark TA, Messonnier NE.** 2011. Current epidemiology and trends in invasive *Haemophilus influenzae* disease—United States, 1989-2008. Clin Infect Dis **53:**1230–1236. http://dx.doi.org/10.1093/cid/cir735.
5. **Murphy TF.** 2015. Vaccines for nontypeable *Haemophilus influenzae*: the future is now. Clin Vaccine Immunol **22:**459–466. http://dx.doi.org/10.1128/CVI.00089-15.
6. **Murphy TF, Brauer AL, Sethi S, Kilian M, Cai X, Lesse AJ.** 2007. *Haemophilus haemolyticus*: a human respiratory tract commensal to be distinguished from *Haemophilus influenzae*. J Infect Dis **195:**81–89. http://dx.doi.org/10.1086/509824.
7. **Chang A, Adlowitz DG, Yellamatty E, Pichichero M.** 2010. *Haemophilus influenzae* outer membrane protein P6 molecular characterization may not differentiate all strains of *H. influenzae* from *H. haemolyticus*. J Clin Microbiol **48:**3756–3757. http://dx.doi.org/10.1128/JCM.01255-10.
8. **Hare KM, Binks MJ, Grimwood K, Chang AB, Leach AJ, Smith-Vaughan H.** 2012. Culture and PCR detection of *Haemophilus influenzae* and *Haemophilus haemolyticus* in Australian Indigenous children with

bronchiectasis. J Clin Microbiol **50:**2444–2445. http://dx.doi.org/10.1128/JCM.00566-12.

9. **Kirkham LA, Wiertsema SP, Mowe EN, Bowman JM, Riley TV, Richmond PC.** 2010. Nasopharyngeal carriage of *Haemophilus haemolyticus* in otitis-prone and healthy children. J Clin Microbiol **48:**2557–2559. http://dx.doi.org/10.1128/JCM.00069-10.
10. **Pickering J, Binks MJ, Beissbarth J, Hare KM, Kirkham LA, Smith-Vaughan H.** 2014. A PCR-high-resolution melt assay for rapid differentiation of nontypeable *Haemophilus influenzae* and *Haemophilus haemolyticus*. J Clin Microbiol **52:**663–667. http://dx.doi.org/10.1128/JCM.02191-13.
11. **Norskov-Lauritsen N.** 2009. Detection of cryptic genospecies misidentified as *Haemophilus influenzae* in routine clinical samples by assessment of marker genes *fucK*, *hap*, and *sodC*. J Clin Microbiol **47:**2590–2592. http://dx.doi.org/10.1128/JCM.00013-09.
12. **Witherden EA, Kunde D, Tristram SG.** 2013. PCR screening for the *N526K* substitution in isolates of *Haemophilus influenzae* and *Haemophilus haemolyticus*. J Antimicrob Chemother **68:**2255–2258. http://dx.doi.org/10.1093/jac/dkt189.
13. **Pickering J, Richmond PC, Kirkham LA.** 2014. Molecular tools for differentiation of non-typeable *Haemophilus influenzae* from *Haemophilus haemolyticus*. Front Microbiol **5:**664. http://dx.doi.org/10.3389/fmicb.2014.00664.
14. **Wang X, Mair R, Hatcher C, Theodore MJ, Edmond K, Wu HM, Harcourt BH, Carvalho MDG, Pimenta F, Nymadawa P, Altantsetseg D, Kirsch M, Satola SW, Cohn A, Messonnier NE, Mayer LW.** 2011. Detection of bacterial pathogens in Mongolia meningitis surveillance with a new real-time PCR assay to detect *Haemophilus influenzae*. Int J Med Microbiol **301:**303–309. http://dx.doi.org/10.1016/j.ijmm.2010.11.004.
15. **Binks MJ, Temple B, Kirkham LA, Wiertsema SP, Dunne EM, Richmond PC, Marsh RL, Leach AJ, Smith-Vaughan HC.** 2012. Molecular surveillance of true nontypeable *Haemophilus influenzae*: an evaluation of PCR screening assays. PLoS One **7:**e34083. http://dx.doi.org/10.1371/journal.pone.0034083.
16. **Hariadi NI, Zhang L, Patel M, Sandstedt SA, Davis GS, Marrs CF, Gilsdorf JR.** 2015. Comparative profile of heme acquisition genes in disease-causing and colonizing nontypeable *Haemophilus influenzae* and *Haemophilus haemolyticus*. J Clin Microbiol **53:**2132–2137. http://dx.doi.org/10.1128/JCM.00345-15.
17. **Theodore MJ, Anderson RD, Wang X, Katz LS, Vuong JT, Bell ME, Juni BA, Lowther SA, Lynfield R, MacNeil JR, Mayer LW.** 2012. Evaluation of new biomarker genes for differentiating *Haemophilus influenzae* from *Haemophilus haemolyticus*. J Clin Microbiol **50:**1422–1424. http://dx.doi.org/10.1128/JCM.06702-11.
18. **Jordan IK, Conley AB, Antonov IV, Arthur RA, Cook ED, Cooper GP, Jones BL, Knipe KM, Lee KJ, Liu X, Mitchell GJ, Pande PR, Petit RA, Qin S, Rajan VN, Sarda S, Sebastian A, Tang S, Thapliyal R, Varghese NJ, Ye T, Katz LS, Wang X, Rowe L, Frace M, Mayer LW.** 2011. Genome sequences for five strains of the emerging pathogen *Haemophilus haemolyticus*. J Bacteriol **193:**5879–5880. http://dx.doi.org/10.1128/JB.05863-11.
19. **Tatusova T, Ciufo S, Federhen S, Fedorov B, McVeigh R, O'Neill K, Tolstoy I, Zaslavsky L.** 2015. Update on RefSeq microbial genomes resources. Nucleic Acids Res **43:**D599–D605. http://dx.doi.org/10.1093/nar/gku1062.
20. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32:**1792–1797. http://dx.doi.org/10.1093/nar/gkh340.
21. **Saitou N, Nei M.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol **4:**406–425.
22. **Kumar S, Stecher G, Tamura K.** 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol http://dx.doi.org/10.1093/molbev/msw054.
23. **Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437:**376–380.
24. **Konstantinidis KT, Tiedje JM.** 2005. Genomic insights that advance the species definition for prokaryotes. Proc Natl Acad Sci U S A **102:**2567–2572. http://dx.doi.org/10.1073/pnas.0409727102.
25. **Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL.** 2004. Versatile and open software for comparing large genomes. Genome Biol **5:**R12. http://dx.doi.org/10.1186/gb-2004-5-2-r12.
26. **Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:**3389–3402. http://dx.doi.org/10.1093/nar/25.17.3389.
27. **Wang X, Cohn A, Comanducci M, Andrew L, Zhao X, MacNeil JR, Schmink S, Muzzi A, Bambini S, Rappuoli R, Pizza M, Murphy E, Hoiseth SK, Jansen KU, Anderson AS, Harrison LH, Clark TA, Messonnier NE, Mayer LW.** 2011. Prevalence and genetic diversity of candidate vaccine antigens among invasive *Neisseria meningitidis* isolates in the United States. Vaccine **29:**4739–4744. http://dx.doi.org/10.1016/j.vaccine.2011.04.092.
28. **Wang X, Theodore MJ, Mair R, Trujillo-Lopez E, du Plessis M, Wolter N, Baughman AL, Hatcher C, Vuong J, Lott L, von Gottberg A, Sacchi C, McDonald JM, Messonnier NE, Mayer LW.** 2012. Clinical validation of multiplex real-time PCR assays for detection of bacterial meningitis pathogens. J Clin Microbiol **50:**702–708. http://dx.doi.org/10.1128/JCM.06087-11.
29. **Mukundan D, Ecevit Z, Patel M, Marrs CF, Gilsdorf JR.** 2007. Pharyngeal colonization dynamics of *Haemophilus influenzae* and *Haemophilus haemolyticus* in healthy adult carriers. J Clin Microbiol **45:**3207–3217. http://dx.doi.org/10.1128/JCM.00492-07.
30. **Norskov-Lauritsen N, Overballe MD, Kilian M.** 2009. Delineation of the species *Haemophilus influenzae* by phenotype, multilocus sequence phylogeny, and detection of marker genes. J Bacteriol **191:**822–831. http://dx.doi.org/10.1128/JB.00782-08.
31. **Pittman M.** 1953. A classification of the hemolytic bacteria of the genus *Haemophilus*: *Haemophilus haemolyticus* Bergey et al. and *Haemophilus parahaemolyticus* nov spec. J Bacteriol **65:**750–751.
32. **Ridderberg W, Fenger MG, Norskov-Lauritsen N.** 2010. *Haemophilus influenzae* may be untypable by the multilocus sequence typing scheme due to a complete deletion of the fucose operon. J Med Microbiol **59:**740–742. http://dx.doi.org/10.1099/jmm.0.018424-0.
33. **Hanage WP.** 2013. Fuzzy species revisited. BMC Biol **11:**41. http://dx.doi.org/10.1186/1741-7007-11-41.
34. **Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, Pati A.** 2015. Microbial species delineation using whole-genome sequences. Nucleic Acids Res **43:**6761–6771. http://dx.doi.org/10.1093/nar/gkv657.
35. **Frickmann H, Christner M, Donat M, Berger A, Essig A, Podbielski A, Hagen RM, Poppert S.** 2013. Rapid discrimination of *Haemophilus influenzae*, *H. parainfluenzae*, and *H. haemolyticus* by fluorescence in situ hybridization (FISH) and two matrix-assisted laser-desorption-ionization time-of-flight mass spectrometry (MALDI-TOF-MS) platforms. PLoS One **8:**e63222. http://dx.doi.org/10.1371/journal.pone.0063222.
36. **Zhu B, Xiao D, Zhang H, Zhang Y, Gao Y, Xu L, Lv J, Wang Y, Zhang J, Shao Z.** 2013. MALDI-TOF MS distinctly differentiates nontypable *Haemophilus influenzae* from *Haemophilus haemolyticus*. PLoS One **8:**e56139. http://dx.doi.org/10.1371/journal.pone.0056139.
37. **Norskov-Lauritsen N, Kilian M.** 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. http://dx.doi.org/10.1099/ijs.0.64207-0.