Genomic Diversity and Antimicrobial Resistance of Haemophilus Colonizing the Airways of Young Children with Cystic Fibrosis

Stephen C. Watts,a,b Louise M. Judd,b Rosemary Carzino,c Sarah Ranganathan,c,d @Kathryn E. Holtab,e

aDepartment of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Melbourne, Victoria, Australia
bDepartment of Infectious Diseases, Central Clinical School, Monash University, Melbourne, Victoria, Australia
cInfection and Immunity, Murdoch Children’s Research Institute, Melbourne, Victoria, Australia
dDepartment of Paediatrics, University of Melbourne, Melbourne, Victoria, Australia
eLondon School of Hygiene & Tropical Medicine, London, United Kingdom

ABSTRACT Respiratory infection during childhood is a key risk factor in early cystic fibrosis (CF) lung disease progression. Haemophilus influenzae and Haemophilus parainfluenzae are routinely isolated from the lungs of children with CF; however, little is known about the frequency and characteristics of Haemophilus colonization in this context. Here, we describe the detection, antimicrobial resistance (AMR), and genome sequencing of H. influenzae and H. parainfluenzae isolated from airway samples of 147 participants aged ≤12 years enrolled in the Australian Respiratory Early Surveillance Team for Cystic Fibrosis (AREST CF) program, Melbourne, Australia. The frequency of colonization per visit was 4.6% for H. influenzae and 32.1% for H. parainfluenzae, 80.3% of participants had H. influenzae and/or H. parainfluenzae detected on at least one visit, and using genomic data, we estimate 15.6% of participants had persistent colonization with the same strain for at least two consecutive visits. Isolates were genetically diverse and AMR was common, with 52% of H. influenzae and 82% of H. parainfluenzae displaying resistance to at least one drug. The genetic basis for AMR could be identified in most cases; putative novel determinants include a new plasmid encoding blaTEM-1 (ampicillin resistance), a new inhibitor-resistant blafop allele (augmentin resistance), and previously unreported mutations in chromosomally carried genes (pdp3, ampicillin resistance; folA/folP, cotrimoxazole resistance; rpoB, rifampicin resistance). Acquired AMR genes were more common in H. parainfluenzae than H. influenzae (51% versus 21%, P = 0.0107) and were mostly associated with the ICEHIn mobile element carrying blafop, resulting in more ampicillin resistance in H. parainfluenzae (73% versus 30%, P = 0.0004). Genomic data identified six potential instances of Haemophilus transmission between participants, of which three involved participants who shared clinic visit days.

IMPORTANCE Cystic fibrosis (CF) lung disease begins during infancy, and acute respiratory infections increase the risk of early disease development and progression. Microbes involved in advanced stages of CF are well characterized, but less is known about early respiratory colonizers. We report the population dynamics and genomic determinants of AMR in two early colonizer species, namely, Haemophilus influenzae and Haemophilus parainfluenzae, collected from a pediatric CF cohort. This investigation also reveals that H. parainfluenzae has a high frequency of AMR carried on mobile elements that may act as a potential reservoir for the emergence and spread of AMR to H. influenzae, which has greater clinical significance as a respiratory pathogen in children. This study provides insight into the evolution of AMR and the colonization of H. influenzae and H. parainfluenzae in a pediatric CF cohort, which will help inform future treatment.

KEYWORDS Haemophilus, Haemophilus influenzae, Haemophilus parainfluenzae, antibiotic resistance, cystic fibrosis, genomics

July/August 2021 Volume 6 Issue 4 e00178-21
Cystic fibrosis (CF) is a common inherited genetic disorder caused by deleterious mutations in the cystic fibrosis transmembrane conductance regulator gene (1). Although the disease is multisystemic, the primary cause of morbidity and mortality results from pulmonary dysfunction. CF lung disease manifests as delayed mucociliary clearance and mucus adhesion leading to recurrent and chronic microbial infections (2), which elicit an adverse host inflammation response resulting in bronchiectasis and ultimately respiratory failure (3, 4). Management of bacterial lung infections is essential in CF disease trajectory and can be managed in part through antimicrobial therapy. However, antimicrobial resistance (AMR) is frequently acquired through various mechanisms and can have clinical consequences in CF patients, including reduced lung function (5–7).

There are a small number of bacterial species that predominantly cause CF lung infections, including *Pseudomonas aeruginosa*, *Burkholderia cepacia* complex, *Staphylococcus aureus*, *Stenotrophomonas maltophilia*, and *Haemophilus influenzae* (8). Importantly, acute respiratory infection in newborns with CF is an established risk factor for early disease development and progression (9).

*H. influenzae* and *Haemophilus parainfluenzae* are among the most common *Haemophilus* species that colonize the respiratory tract of children in early life (10, 11). *H. influenzae* is considered an opportunistic pathogen and can cause invasive disease. Instances of invasive *H. parainfluenzae* infection have also been described (12–14); however, *H. parainfluenzae*-related disease is less frequently observed, and *H. parainfluenzae* is recognized as having a lower pathogenic capacity than *H. influenzae*. Both *H. influenzae* and *H. parainfluenzae* are routinely isolated from the respiratory tract of children with CF, particularly during episodes of disease exacerbation (15). Although many of the classical pathogens involved in CF lung disease have been well studied, less is known about the role of *Haemophilus* species during the critical period of early childhood. Such knowledge is essential, as it is increasingly recognized that CF lung disease commences soon after diagnosis in early infancy and progresses thereafter (16, 17). Further insight into the epidemiology and resistance profiles of these early colonizing and infecting bacteria will inform future treatment practices.

The emergence and accumulation of AMR in *H. influenzae* and *H. parainfluenzae* is common, with the highest resistance rates reported for ampicillin (AMP; 23.9 to 58.5% in *H. influenzae*, 13.2 to 50.0% in *H. parainfluenzae*) and cotrimoxazole (21.4 to 71.1% in *H. influenzae*, 14.9 to 44.2% in *H. parainfluenzae*) (18–31). Generally, *H. influenzae* infections are treated with β-lactams, such as extended-spectrum penicillins or cephalosporins (32). Other drugs are often used in combination with or as an alternative to β-lactams and include antifolates, quinolones, and macrolides. Resistance to these drugs typically arises through either acquisition of horizontally transferred resistance genes or mutations in chromosomally encoded protein targets (32). Acquired AMR genes in *H. influenzae* and *H. parainfluenzae* are frequently localized within mobile genetic elements, such as ICEHin or small plasmids (33–35), which appear to have facilitated the emergence of multidrug-resistant *H. influenzae* and *H. parainfluenzae* strains in recent years (36, 37).

*H. influenzae* is known to produce a polysaccharide capsule, which can be classified into six serotypes (Hia through Hif) and is an invasive virulence determinant (38). Strains that do not produce the capsule are designated nontypeable *H. influenzae* (NTHi). The introduction of the highly effective Hib conjugate vaccines caused a marked reduction of Hib-related disease incidence but consequently resulted in an increased prevalence of NTHi-related disease (39); NTHi is now more commonly isolated from children with CF than any encapsulated *H. influenzae* serotype (40, 41). *H. parainfluenzae* is generally less well characterized, and the role it may have in CF disease is unclear. There is no detailed description of encapsulated *H. parainfluenzae*, although there is increasing evidence that some *H. parainfluenzae* strains could express a polysaccharide capsule (37). Moreover, there is a stark lack of *H. parainfluenzae* genomic data compared with *H. influenzae*, despite that it occupies a similar niche. Here, we investigate the prevalence, genomic diversity, and AMR phenotypes of *H. influenzae* and *H. parainfluenzae* colonizing the airways of children with CF, recruited at the Royal Children’s Hospital (RCH), Melbourne, Australia.
TABLE 1 Study participant characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>All participants (lab-reported species identification)</th>
<th>Participants with WGS data (WGS species identification)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of total participants</td>
<td>147</td>
<td>59</td>
</tr>
<tr>
<td>No. of females (%)</td>
<td>64 (43.5)</td>
<td>23 (38.9)</td>
</tr>
<tr>
<td>Mean age at first sample, yrs (range)</td>
<td>5.7 (0.08–11.8)</td>
<td>4.1 (0.10–8.9)</td>
</tr>
<tr>
<td>Mean no. of samples, count (range)</td>
<td>5.8 (1–12)</td>
<td>6.2 (2–12)</td>
</tr>
<tr>
<td>No. with ≥1 H. influenzae-positive sample (%)</td>
<td>30 (20.4)</td>
<td>21 (35.6)</td>
</tr>
<tr>
<td>No. with ≥1 H. parainfluenzae-positive sample (%)</td>
<td>111 (73.5)</td>
<td>55 (93.2)</td>
</tr>
<tr>
<td>No. with Zero H. influenzae- or H. parainfluenzae-positive samples (%)</td>
<td>29 (19.7)</td>
<td></td>
</tr>
</tbody>
</table>

Participants with ≥1 Haemophilus culture-positive sample

| No. of total participants | 118                                                   | 59                                                     |
| No. of females (%) | 54 (45.8)                                             | 23 (39.0)                                              |
| Mean age at first sample, yrs (range) | 5.3 (0.08–11.8)                                  | 4.1 (0.10–8.9)                                        |
| Mean no. of samples (range) | 6.0 (2–12)                                          | 6.2 (2–12)                                             |
| Mean no. of H. influenzae-positive samples (range) | 0.3 (0–3)                                         | 0.5 (0–3)                                              |
| Mean no. of H. parainfluenzae-positive samples (range) | 2.3 (0–7)                                         | 2.8 (0–7)                                              |

RESULTS

Detection and sequencing of Haemophilus isolates. During the 1-year study period, 147 AREST CF study participants receiving treatment at the RCH CF specialist clinic were screened for the presence of H. influenzae or H. parainfluenzae in the respiratory tract during regular clinic visits and during hospitalization for pulmonary exacerbation. Participant characteristics are given in Table 1. Thirty (20.4%) participants tested culture positive for colonization with H. influenzae in ≥1 sample and 111 (75.5%) participants for H. parainfluenzae. Only 29 participants (19.7%) had no Haemophilus-positive samples; these individuals did not differ by age or gender but contributed fewer samples each (mean 4.6 versus 6.0, P = 0.012 using two-sample Kolmogorov-Smirnov test). The overall frequency of colonization per visit was 4.6% for H. influenzae and 32.1% for H. parainfluenzae, with 86 participants (58.5%) presenting with either H. influenzae or H. parainfluenzae on 2 or more occasions. Several participants had both H. influenzae and H. parainfluenzae detected during the same (n = 10, 6.8%) or different (n = 15, 10.2%) visits.

A subset of 162 Haemophilus isolates from 59 participants (50% of culture-positive individuals, representative in terms of age and gender) (see Table 1) were subjected to whole-genome sequencing (WGS), of which 89.5% were sequenced successfully. WGS data revealed some species misidentifications (5.5%) and mixed cultures (13%), leaving 24 H. influenzae and H. parainfluenzae isolates for further analysis (see Fig. S1 online at doi.org/10.26188/14954931 and Materials and Methods).

Genetic diversity and population structure. The H. influenzae isolates collected in this study were highly diverse (median of 2.3% nucleotide divergence in conserved core genome), and a comparison with publicly available WGS data from other studies (n = 877 genomes, see Data Set S1 in the supplemental material) indicates RCH CF isolates are distributed across the global species phylogeny (red tips in Fig. 1A). All but two RCH CF H. influenzae genomes had no detectable capsule biosynthesis (cap) locus and fell outside the small number of lineages typically associated with encapsulation (42) (colored branches in Fig. 1A); they are thus predicted to be unencapsulated or nontypeable. Two RCH CF isolates (drug susceptible, sequence type 18 [ST18], from the same patient) carried intact copies of the cap-e locus and fell within the lineage typically associated with serotype e (green branches in Fig. 1A); thus, they are predicted to express serotype e capsules.

We used discriminant analysis of principal components (DAPC) to explore how the population of RCH CF H. influenzae isolates compared with the global H. influenzae population structure captured by k-mer profiles of publicly available WGS data from a range of other contexts (see Data Set S1; see Fig. S4 online at doi.org/10.26188/14954931).
data indicate that our Australian pediatric CF isolates are typical of noninvasive respiratory isolates from children in other settings, based on functions constructed to discriminate these parameters from isolates collected from adults, nonrespiratory specimens, and invasive disease (Fig. 2A to C). RCH isolates clustered most closely with other Australian isolates in the discriminant function based on geographical location (Fig. 2D). Notably, DAPC analysis of the 264,940 core-genome single nucleotide variants (SNVs) used for phylogenetic inference yielded much weaker discriminant functions for specimen type and geographical location (see Fig. S5 online at doi.org/10.26188/14954931), suggesting that the discriminant genetic features are carried in the accessory genome rather than core gene allelic variation.

The *H. parainfluenzae* isolates collected in this study also show extensive genetic diversity (median 5.1% nucleotide divergence in conserved core genome). This finding is
harder to contextualize within the overall species diversity due to the low number of genomes available from other studies but appears to be representative of species-wide phylogenetic diversity (Fig. 1B).

**Antimicrobial resistance.** AMR was relatively common, with 52% of *H. influenzae* and 82% of *H. parainfluenzae* isolates displaying resistance to at least one of the five drugs tested (Table 2). The frequency of cotrimoxazole (STX) resistance was similar in both species (35% in *H. influenzae*, 31% in *H. parainfluenzae*), but ampicillin and/or augmentin (AMC) resistance was observed at significantly higher rates in *H. parainfluenzae* than those in *H. influenzae* (P < 0.05) (see Table 2). Rifampicin (RIF) was observed at low frequencies in both species (17% in *H. influenzae*, 7% in *H. parainfluenzae*). Multidrug resistance (defined here as resistance to ampicillin or augmentin plus at least one other drug class) was also more commonly detected in *H. parainfluenzae*, although the difference was not statistically significant (see Table 2).

We used the WGS data to explore genetic determinants of AMR in the RCH isolates. Horizontally acquired AMR genes were more frequently found in *H. parainfluenzae* than in *H. influenzae*, with 42 (51%) *H. parainfluenzae* and 5 (21%) *H. influenzae* isolates containing one or more acquired AMR genes (P = 0.0107 using Fisher’s exact test) (see DATA SET S2 in the supplemental material). Most common were *bla*<sub>TEM</sub> genes (36 *bla*<sub>TEM-1</sub>, 4 *bla*<sub>TEM-40</sub>, and 2 *bla*<sub>TEM-30</sub>), carried in the mobile element ICE<sub>HiN</sub> (2 *H. influenzae*, 38 *H. parainfluenzae*) or small plasmids (3 *H. influenzae*, 2 *H. parainfluenzae*). Other AMR genes were less common and restricted to *H. parainfluenzae*, which were generally located in ICE<sub>HiN</sub> elements (11 isolates carried *strAB* [aminoglycosides] and *sul1* [sulfonamide-cotrimoxazole], 5 *aph*<sub>3’</sub>*a* [aminoglycosides], 1 *tetB* [tetracycline], and 1 *tetM* [tetracycline] with *msrD* and *mefA* [macrolides]). The resistance cassettes of ICE<sub>HiN</sub> varied in structure and gene content (Fig. 3A). Four distinct *bla*<sub>TEM</sub> plasmids were observed and were of a similar size (4.3 to 6.5 kbp). Three plasmids were homologous to previously sequenced *H. influenzae* plasmids.
pLFH64 (2 H. influenzae, blaTEM-1), pA1209 (one H. influenzae, blaTEM-1), pPN223 (one H. parainfluenzae, blaTEM-1); the fourth was a novel plasmid present in one H. parainfluenzae isolate (pM1C124_1, blaTEM-40; deposited in GenBank under accession MW111541) (Fig. 3B; see Table S4 online at doi.org/10.26188/14954931). Acquired AMR genes accounted for only 33.6% of observed nonsusceptibility phenotypes; hence, we screened for mutations in conserved core resistance-related chromosomal genes that could potentially explain resistance to ampicillin and augmentin (pbp genes) (see Table S3 online at doi.org/10.26188/14954931), cotrimoxazole (folP and folA), and rifampicin (rpoB) (summarized in Tables S5 and S6 at doi.org/10.26188/14954931).

Ampicillin resistance in H. influenzae could be entirely explained by acquired β-lactamases encoded by blaTEM genes (57%) and/or mutations in the penicillin-binding proteins PBP3 (FtsI) and PBP1B (MrcB) (Table S5 online at doi.org/10.26188/14954931). In H. parainfluenzae, acquired blaTEM could explain 60% of ampicillin resistance, but we detected no known or novel PBP mutation that was statistically associated with resistance (Table S6 online at doi.org/10.26188/14954931). Notably though, the first H. parainfluenzae isolated from participant M1C152 was ampicillin sensitive and wild type at PBP3-502. The two subsequent H. parainfluenzae isolates from this participant (following treatment with augmentin and ceftriaxone) were ampicillin resistant with no acquired AMR genes and differed from the

**TABLE 2** Frequency of nonsusceptibility to antimicrobials among sequenced isolates

<table>
<thead>
<tr>
<th>Antimicrobial</th>
<th>H. influenzae</th>
<th>H. parainfluenzae</th>
<th>Odds ratio (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ampicillin</td>
<td>7/23 (30)</td>
<td>60/82 (73)</td>
<td>6.11 (2.05–20.09)</td>
<td>0.0004</td>
</tr>
<tr>
<td>Augmentin</td>
<td>1/17 (6)</td>
<td>25/79 (32)</td>
<td>7.30 (1.02–322.38)</td>
<td>0.0351</td>
</tr>
<tr>
<td>Cefotaxime</td>
<td>0/23 (0)</td>
<td>2/81 (3)</td>
<td>Inf (0.053–Inf)</td>
<td>1</td>
</tr>
<tr>
<td>Cotrimoxazole</td>
<td>8/23 (35)</td>
<td>25/82 (31)</td>
<td>0.82 (0.28–2.55)</td>
<td>0.8002</td>
</tr>
<tr>
<td>Rifampicin</td>
<td>2/12 (17)</td>
<td>4/59 (7)</td>
<td>0.37 (0.045–4.61)</td>
<td>0.2659</td>
</tr>
<tr>
<td>Ampicillin or augmentin</td>
<td>7/17 (41)</td>
<td>62/78 (80)</td>
<td>5.41 (1.58–19.72)</td>
<td>0.0026</td>
</tr>
<tr>
<td>One or more drugs tested</td>
<td>12/23 (52)</td>
<td>68/83 (82)</td>
<td>4.09 (1.36–12.47)</td>
<td>0.0058</td>
</tr>
<tr>
<td>MDR</td>
<td>5/23 (22)</td>
<td>23/83 (28)</td>
<td>1.38 (0.42–5.30)</td>
<td>0.7897</td>
</tr>
</tbody>
</table>

Nonsusceptibility was defined as I or R according to clinical breakpoints (see Materials and Methods). Multidrug resistance (MDR) was defined as resistance to ampicillin or augmentin plus at least one other antimicrobial. Association tests compare resistance rate between species and were performed using Fisher’s exact test. The analysis is restricted to isolates that were successfully sequenced, had a susceptibility phenotype reported for the given antimicrobial(s), and found to be pure cultures, with species identification based on genome data. Inf, infinity.

*pLFH64 (2 H. influenzae, blaTEM-1), pA1209 (one H. influenzae, blaTEM-1), pPN223 (one H. parainfluenzae, blaTEM-1); the fourth was a novel plasmid present in one H. parainfluenzae isolate (pM1C124_1, blaTEM-40; deposited in GenBank under accession MW111541) (Fig. 3B; see Table S4 online at doi.org/10.26188/14954931). Acquired AMR genes accounted for only 33.6% of observed nonsusceptibility phenotypes; hence, we screened for mutations in conserved core resistance-related chromosomal genes that could potentially explain resistance to ampicillin and augmentin (pbp genes) (see Table S3 online at doi.org/10.26188/14954931), cotrimoxazole (folP and folA), and rifampicin (rpoB) (summarized in Tables S5 and S6 at doi.org/10.26188/14954931).

Ampicillin resistance in H. influenzae could be entirely explained by acquired β-lactamases encoded by blaTEM genes (57%) and/or mutations in the penicillin-binding proteins PBP3 (FtsI) and PBP1B (MrcB) (Table S5 online at doi.org/10.26188/14954931). In H. parainfluenzae, acquired blaTEM could explain 60% of ampicillin resistance, but we detected no known or novel PBP mutation that was statistically associated with resistance (Table S6 online at doi.org/10.26188/14954931). Notably though, the first H. parainfluenzae isolated from participant M1C152 was ampicillin sensitive and wild type at PBP3-502. The two subsequent H. parainfluenzae isolates from this participant (following treatment with augmentin and ceftriaxone) were ampicillin resistant with no acquired AMR genes and differed from the

**Fig 3** Representative ICEHin structures (A) and plasmids (B) carrying AMR genes identified in RCH isolates. Plasmids are annotated with the best corresponding match in the NCBI nucleotide database, see Table S3 online at doi.org/10.26188/14954931.
first by a single SNV across the entire genome resulting in the amino acid substitution PBP3-A502T, supporting the previously reported role of this mutation in conferring resistance (43). Nevertheless, our findings leave 33% of ampicillin resistance in \textit{H. parainfluenzae} unexplained.

Augmentin is a combination of amoxicillin plus the \( \beta \)-lactamase inhibitor clavulanic acid. Just four amoxicillin-resistant isolates (16%) carried inhibitor-resistant \( \beta \)-lactamase alleles (\( 2 \text{ H. parainfluenzae with bla}_{\text{TEM-30}} \), \textit{2 H. parainfluenzae with bla}_{\text{TEM-40}}). Nine more isolates (36%) carried \( \text{bla}_{\text{TEM-1r}} \), but this encoded \( \beta \)-lactamase is susceptible to clavulanic acid inhibition and we identified no \( \text{pbp} \) variants in these isolates that were significantly associated with augmentin resistance; hence, inhibitor resistance remains unexplained in these cases. Two amoxicillin-resistant \( \text{H. parainfluenzae} \) isolates collected from the same participant (M1C141) contained novel \( \text{bla}_{\text{TEM}} \) alleles that share substitution mutations with known inhibitor-resistant alleles (\( \text{bla}_{\text{TEM-1r}} \)-M671, \( \text{bla}_{\text{TEM-1r}} \\ W163L) (Fig. S6 online at doi.org/10.26188/14954931), which likely explain the phenotype (44, 45). Hence, the vast majority of augmentin resistance (100% in \textit{H. influenzae}, 80% in \textit{H. parainfluenzae}) is unexplained. Resistance to the third-generation cephalosporin cefotaxime (CTX) was observed in two \textit{H. parainfluenzae} isolates from different participants but was unexplained (one carried no acquired genes, one carried only \( \text{bla}_{\text{TEM-1r}} \), and neither carried unique \( \text{pbp} \) mutations).

Cotrimoxazole is a combination of trimethoprim and sulfamethoxazole. Resistance to trimethoprim is associated with mutations in the chromosomal dihydrofolate reductase \( \text{folA} \) or acquisition of mobile resistant alleles (\( \text{dfr} \) genes), while resistance to sulfamethoxazole requires mutations in the chromosomal dihydropteroate synthase \( \text{folP} \) or acquisition of mobile resistant alleles (\( \text{sul} \) genes). In \textit{H. influenzae}, no acquired \( \text{sul} \) or \( \text{dfr} \) genes were detected; however, all cotrimoxazole-resistant isolates carried a novel resistance-associated mutation, \( \text{FoIA-N13S} \), and most carried the novel \( \text{FoIP-G189C} \) (75%), as well as previously reported \( \text{FoIA-I95L} \) (75%) and \( \text{FoIP-P64ins} \) (38%) (Table S5 online at doi.org/10.26188/14954931, and Data Set S2). In \textit{H. parainfluenzae}, 92% of cotrimoxazole-resistant isolates carried \( \text{sulI} \) (36%) and/or resistance-associated \( \text{FoIP} \) mutations (including \( \text{FoIP-P64ins} \) and \( \text{FoIP-G189C} \), 80%); 64% carried resistance-associated \( \text{FoIA} \) mutations (46) (Table S6 online at doi.org/10.26188/14954931, and Data Set S2).

Rifampicin resistance is most often explained by mutations in \( \text{rpoB} \) (the RNA polymerase beta subunit), including one previous report in \textit{H. influenzae} (47). Both rifampicin-resistant \textit{H. influenzae} isolates (from the same patient, M1C073) carried a novel mutation, \( \text{RpoB-A1131T} \), that was absent from sensitive isolates. In \textit{H. parainfluenzae}, one of four rifampicin-resistant isolates (M1C081_2) carried \( \text{RpoB-T724I} \), which has been previously described in resistant \textit{H. parainfluenzae} isolates (19); however, the remaining three isolates contained no other mutations associated with rifampicin resistance (Data Set S2).

\textbf{Persistent colonization and transmission.} Seventy-nine participants (53.7%) were culture positive for the same \textit{Haemophilus} species on \( \geq 1 \) occasion; 7 (4.8%) participants had \( \geq 2 \) \textit{H. influenzae} and 75 (51%) had \( \geq 2 \) \textit{H. parainfluenzae}. The probability of testing culture positive for the same species in the next sample after an initial positive result (mean time interval 105 days) was 16.0% for \textit{H. influenzae} and 48.1% for \textit{H. parainfluenzae} (\( P = 0.004 \) for test of difference in proportions). Among those individuals who had a culture-positive sample directly followed by a culture-negative for the same species (\( n = 107 \)), 36 (33.6%) had a subsequent positive sample and 11 (10.3%) had no further samples tested. In 45/79 participants, WGS data were available for at least 2 isolates of the same species. Among these patients with \( \geq 2 \) WGS sequences, 13 participants (29%) had matching isolates of the same strain (defined as \( \leq 20 \) mutations; see Materials and Methods), consistent with persistent colonization (2 \textit{H. influenzae} and \( 11 \textit{H. parainfluenzae} \) (see Fig. 4A). Assuming the same rate of strain matching (29%) among the 34 participants who had \( \geq 2 \) isolates but WGS data was available for only 1 of those isolates, we estimate that a further 10 of these participants would have matching strains. Thus, we estimate a total of 23 (15.6%) of the 147 participants (95% confidence interval [CI], 8.4% to 22.6%) had \textit{Haemophilus} colonization that persisted between visits. Notably,
the only encapsulated strain detected in this study (cap-e H. influenzae) was detected twice in the same participant (M1C094), on two separate clinic visits 84 days apart.

WGS data showed that most strains were unique to a single participant; however, we identified 6 participant pairs that shared the same strain (1 to 19 mutations, see Materials and Methods) and those below are different strains. Dist, range of pairwise genetic distances (nonrecombinant SNVs + number of inferred recombination events) observed between isolates of the same strain. In B, lines connecting participant timelines represent instances where participants attended an RCH clinic during the same day. Shared days, number of days on which both participants attended the RCH CF clinic; 1° intermediaries, for participant pairs not sharing any clinic visit days with one another, we searched for primary intermediary participants who shared at least one clinic data point with each of the participants.

Twelve of the 13 (92%) WGS-confirmed cases of persistent strain colonization exhibited resistance to at least 1 drug, compared with 70% of strains that were not identified as persisting (see Table S7 online at doi.org/10.26188/14954931) (Data Set S2). For H. parainfluenzae, all
AMR phenotypes were more frequent among isolates associated with persistent strain colonization; however, these comparisons were underpowered, and the differences were only statistically significant for ampicillin, augmentin, and cefotaxime (Table S7 at doi.org/10.26188/14954931). Five of the six potentially transmitted strains displayed at least one AMR phenotype, similar to the overall rate of AMR across all colonizing isolates (Table S8 online at doi.org/10.26188/14954931) (Data Set S2). Changes in AMR phenotypes within individual *H. influenzae* or *H. parainfluenzae* strains were observed during both persistent colonization (resistance phenotypes varied in 8/13 individuals, 62%) and transmission chains (resistance phenotypes varied in 4/6 transmission pairs, 67%), indicating short-term evolution of resistance (Data Set S2).

**DISCUSSION**

*H. influenzae* and *H. parainfluenzae* colonization of the airways was strikingly common in this cohort, with $\geq 80\%$ of participants contributing $\geq 1$ *Haemophilus* culture-positive respiratory sample during the 1-year period of study (Table 1) and 58.5% contributing $\geq 2$ such samples. The point prevalence was 4.6% for *H. influenzae* and 32.1% for *H. parainfluenzae*, and repeat colonization with *H. parainfluenzae* was much more common than that with *H. influenzae* (detected in 51% and 4.8% of participants, respectively). *H. parainfluenzae* also displayed a significantly higher frequency of AMR (Table 2), which was perhaps linked to its increased carriage rate. Globally, the *H. influenzae* carriage rate in children varies, likely due to differences in cohort demographics and geographical location. *H. influenzae* has been reported to be recoverable from the nasopharynx in 8% to 34% of children with CF (40, 48–52), consistent with *H. influenzae* carriage estimation in this CF cohort. The frequency of *H. parainfluenzae* airway colonization has not been established in children (with or without CF) despite the potential to opportunistically cause disease and act as a reservoir for AMR genes.

Substantial genetic diversity was observed for both *H. influenzae* and *H. parainfluenzae* isolates cultured from the airways of participants in this study (Fig. 1). Unsurprisingly, only two *H. influenzae* isolates (from a single patient) were predicted to be encapsulated (cap-e); they belonged to a known clonal capsule-positive lineage (ST18). The remaining *H. influenzae* isolates belong to the highly heterogenous NTHi group, similar to those detected in other studies examining nasopharyngeal colonization, which consistently report NTHi as the dominant *H. influenzae* subtype in the respiratory tract (40, 41).

An analysis of *H. influenzae* core-genome SNVs using phylogenetics and DAPC showed no apparent lineage associations with age group, specimen type, disease status, or geographical location (Fig. S4 and S5 online at doi.org/10.26188/14954931). This finding is consistent with prior studies of NTHi which reported finding no evidence for phylogenetic signals of geographical origin (53, 54) or clinical source (54, 55). However, whole-genome k-mer DAPC revealed distinct clustering of RCH CF isolates with others that were noninvasive, collected from the respiratory tract, isolated from children, and circulating in Australia (based on the respective individual discriminant functions, see Fig. 2). Hence, *H. influenzae* isolates of distinct epidemiological origins are differentiable based on variation in accessory genes but not by allelic variation in the core genome.

The structured variability across the accessory genome could potentially be explained by niche-specific positive selection of genes that confer increased fitness. For example, fixation of ICE*Hin*1056 in respiratory *H. influenzae* populations has been previously observed within 2 weeks of amoxicillin treatment but subsequently lost (or resistant strains outcompeted) 12 weeks after the initial treatment (56). The Hla and HMW adhesins, Hif pilus, and IgA proteases are *H. influenzae* virulence factors that are also differentially present in strains (53, 57) and play a role in the colonization of specific niches like the respiratory tract (58–61). Genome-wide association analysis could potentially identify other contributing factors (62); however, this identification is beyond the scope of the present study.

Antimicrobial therapy is used routinely both to control bacterial lung infections of CF patients (63) and also as an antimicrobial prophylaxis. Augmentin is routinely used for both of these purposes. Cotrimoxazole is used at many specialist CF centers, but it is not
routine at RCH; however, resistance was still observed in nearly a one-third of *H. influenzae* and *H. parainfluenzae*. Regular use of antimicrobials is known to induce resistance, and indeed, we observed a high rate of AMR in isolates collected in this study, with resistance to one or more drugs observed in 52% *H. influenzae* and 82% *H. parainfluenzae* (Table 2). AMR rates in *H. influenzae* isolated from the respiratory tract in non-CF patients vary between studies, with recent reports of ampicillin resistance at 23.9% to 58.5%, augmentin at 0% to 10.4%, cefotaxime at 0% to 5.9%, cotrimoxazole at 51.2% to 71.1%, and rifampicin at 0% and 4.8% (18–27). Similar rates have been reported for *H. parainfluenzae* in non-CF patients, as follows: for ampicillin, 13.2% to 18.5%; augmentin, 0% to 12.5%; cefotaxime, 0% to 0.3%; cotrimoxazole, 14.9% to 44.2%; and rifampicin, 26.7% (28–30). AMR rates detected in the present study are mostly in line with rates in these reports, with the exception of higher rates of resistance in *H. parainfluenzae* versus *H. influenzae*. The only other report of such a difference is an earlier study in our setting (children with CF at RCH, 1998 to 2012) (31), which also found higher rates of resistance in *H. parainfluenzae* than in *H. influenzae* and showed that rates of ampicillin, augmentin, and cotrimoxazole resistance increased significantly in *H. parainfluenzae* over the 15-year duration of the study.

Most of the AMR phenotypes were explained by the presence of known genetic determinants. Ampicillin resistance was the most readily explained by known mechanisms, with all *H. influenzae*-resistant isolates and 67% of *H. parainfluenzae*-resistant isolates harboring the acquired gene *bla*<sub>TEM</sub> or resistance-associated mutations in PBP genes (Data Set S2). The exceptions were augmentin and ceftriaxone; inhibitor-resistant *bla*<sub>TEM</sub> alleles accounted for just 20% of augmentin resistance in *H. parainfluenzae* and none in *H. influenzae*, and no mechanisms for ceftriaxone resistance were identified.

Novel mutations in AMR-associated proteins discovered through association analysis increased the proportion of resistance explained by amino acid substitutions from 13.4% to 31.3%. Several mutations in FolA and FolP were associated with cotrimoxazole resistance (Table S5 and S6 online at doi.org/10.26188/14954931), including both novel and previously established mutations (46). Notably, we identified an insertion in *H. parainfluenzae* FolP that was strongly linked with cotrimoxazole resistance and shared the same location as the *H. influenzae* FolP-P64ins mutation, which has been demonstrated to induce sulfamethoxazole resistance (64). Mutations for rifampicin resistance were identified in *H. influenzae* (RpoB-A1131T) and *H. parainfluenzae* (RpoBT724I); the latter was also recently reported as resistance associated in an independent study of *H. parainfluenzae* (19).

Consistent with previous studies, nearly all acquired genes detected here in *Haemophilus* isolates were localized to either an ICE<sub>Hin</sub> element (65, 66) or small *bla*<sub>TEM</sub> plasmids (34, 44, 67). Novel variants of acquired resistant determinants were also observed, including a new ICE<sub>Hin</sub>-encoded *bla*<sub>TEM</sub> allele associated with augmentin resistance in *H. parainfluenzae* and a novel plasmid harboring *bla*<sub>TEM</sub>-1.

Not all AMR could be explained by an underlying genetic component. This result is likely due in part to a lack of statistical power for detecting novel resistance-associated variants, even when taking a candidate-gene approach as we did, due to the small sample size. This limitation is particularly problematic for *H. influenzae*, for which only 24 sequenced isolates were available; for example, FolP-G189C was associated with cotrimoxazole resistance in both *H. influenzae* and *H. parainfluenzae* but was statistically significant only in *H. parainfluenzae* after adjustment for multiple testing (Table S5 and S6 online at doi.org/10.26188/14954931). Additionally, AMR phenotypes are not always reproducible, and AMR genes or mutations can be lost during subculture to extract DNA for sequencing. Moreover, it is conceivable that some single chromosomal mutations reported here are alone insufficient to confer resistance and instead may require a stepwise acquisition of additional mutations before resistance is gained.

A small number of isolates originally identified biochemically as *H. influenzae* were found to be *H. parainfluenzae* via WGS (n = 3) and vice versa (n = 5; Fig. S1 online at doi.org/10.26188/14954931). The definitive underlying cause of this discrepancy is unclear but could be explained by several possibilities, including the presence of both *H. influenzae* and *H. parainfluenzae*.
and *H. parainfluenzae* in the same sample or inaccuracies in the biochemical species identification test. The overall rate of discordance between biochemical and genomic species identification was $\leq 7\%$, suggesting that studies of *H. influenzae* colonization or infection that rely solely on biochemical identification without additional confirmation may suffer from both false positives and false negatives. There is little published data on the persistence of *Haemophilus* colonization in the lungs of children; however, there is some evidence that *H. influenzae* strain persistence is associated with chronic respiratory disease and does not occur in healthy childhood cohorts (68). This study reveals for the first time strain persistence of *H. influenzae* and *H. parainfluenzae* in the lungs of children with cystic fibrosis for up to 349 days and estimates the carriage rate of persistent strains in the cohort to be 15.6% (95% CI, 8.4% to 22.6%). Strain persistence likely arises due to substantial selective pressure exerted by extensive and prolonged administration of antimicrobials or niche adaptation to the diseased lung. For example, mutations in the single-strand mispairing mechanism allow *H. influenzae* to alter the expression of nutrient uptake systems and surface molecules, such as adhesins, during persistent colonization in adult patients with chronic obstructive pulmonary disorder (54). Other important CF pathogens, such as *P. aeruginosa* and members of the *Burkholderia cepacia* complex, have been shown to undergo similar changes to surface molecules and remodeling of regulatory networks during persistence (69–71). In addition to strain persistence, we observed that participants were frequently colonized by different strains of *H. influenzae* and *H. parainfluenzae* across clinic visits, indicating that *Haemophilus* colonization is a dynamic process and suggesting that strains of both species compete to occupy the niche.

There were six instances where participants shared the same *Haemophilus* strain. Three of the six cases were supported by epidemiological links whereby participants shared clinic visit days (Fig. S7 online at doi.org/10.26188/14954931), and the remaining three cases shared visit days with possible intermediaries (Fig. S7 online at doi.org/10.26188/14954931). Nosocomial transmission of CF pathogens, such as *P. aeruginosa*, has been demonstrated in other settings (71, 72) and historically at our center (73, 74), as has cross-infection with *Mycobacterium abscessus* (75). These findings have led to strict infection control practices in CF clinics such as RCH with strict isolation in both clinics and inpatient areas, wearing of face masks by patients in all public spaces, and strict gloving and gowning by all clinical staff. Notably, preceding the introduction of such stringent infection-control measures, sharing of RCH CF clinic visit days was common in our cohort, and nearly all participant pairs could be connected either through a shared clinic visit day (15%) or shared visit days with a single intermediary (72.4%). Hence, it is not clear whether the overlap in visit days could be circumstantial or strain sharing may reflect circulation of strains in the general community rather than nosocomial transmission. Future studies in settings where fewer patients share visit days may be better able to differentiate these possibilities.

This study provides the first insights into the population dynamics and genomic determinants of AMR among colonizing *H. influenzae* and *H. parainfluenzae* strains in a pediatric CF cohort and identifies multiple novel AMR determinants particularly for *H. parainfluenzae*. Notably, while relatively little attention has been paid to *H. parainfluenzae* colonization in children due to its relative lack of pathogenicity, our data indicate it is a common colonizer that can persist in the respiratory tract of CF children and is very frequently drug resistant. The high frequency of AMR in *H. parainfluenzae*, of which most was encoded in mobile elements that can transfer to *H. influenzae*, indicates that *H. parainfluenzae* could serve as a reservoir for the emergence and spread of AMR to *H. influenzae* which is of more significant clinical concern in children with and without CF. Further insights are essential and will inform antimicrobial treatment and stewardship in the future. Understanding the role of *H. influenzae* and *H. parainfluenzae* in early CF disease progression falls within the province of the AREST CF program goals, and additional studies will aim to assess and explore the specific risk factors associated with early lung colonization by these *Haemophilus* species.

**MATERIALS AND METHODS**

**Participant recruitment and sample and data collection.** Participants in this study are a subset of those enrolled in the Australian Respiratory Early Surveillance Team for Cystic Fibrosis (AREST CF) birth...
cohort who meet the following inclusion criteria: diagnosed with CF, under 12 years of age, resident in catchment area, and presented to the RCH CF clinic between February 2016 to February 2017 (16). Respiratory samples (bronchoalveolar lavage [BAL] fluid, sputum, or cough swabs) were routinely collected from participants during regular visits and cultured on chocolate agar in the RCH microbiological diagnostics laboratory as previously described (16). During the 1-year study period, 847 samples collected from 147 study participants were analyzed and yielded 39 isolates identified as *H. influenzae* and 272 identified as *H. parainfluenzae* (identified using the X and V factor test). Isolates were tested for susceptibility to ampicillin (AMP), augmentin (AMC), cefotaxime (CTX), cotrimoxazole (STX), and rifampicin (RIF) using disk diffusion with CLSI breakpoints (Table S1 online at doi.org/10.26188/14954931).

**Bacterial isolates, sequencing, and assembly.** A total of 162 of the 311 *Haemophilus* isolates (*n* = 30, 77% of those biochemically identified as *H. influenzae; and *n* = 132, 48.5% as *H. parainfluenzae*) were successfully resuscitated, subcultured, and transferred to the University of Melbourne for whole-genome sequencing (WGS). Isolates were plated onto chocolate agar and incubated at 37°C under microaerophilic conditions for 48 hours. Colonies were harvested and DNA extracted using GenFindV2 (Beckman Coulter), using proteinase K for bacterial lysis according to the manufacturer’s instructions. Short-read DNA libraries were prepared for all isolates with a Nextera XT kit (Illumina) and subsequently sequenced on the Illumina MiniSeq platform, generating paired-end reads of 151 bp each. DNA samples for long-read sequencing were prepared for a subset of 14 isolates using GenFindV2 (Beckman Coulter); a barcoded ligation library was prepared (SQK-LSK108, EXP-NBD103) and sequenced via an Oxford Nanopore MinION device on a R9.4.1 flow cell.

A total of 107 isolates (24 *H. influenzae*, 83 *H. parainfluenzae*) were successfully sequenced via Illumina and passed quality control, each yielding ≥150,000 high-quality reads (Fig. S1 online at doi.org/10.26188/14954931 and Data Set S1). Centrifuga v1.0.4b (76) was used to categorize isolates as either (i) pure *H. influenzae* or *H. parainfluenzae* defined as one of these species at ≥50% relative abundance and the next most common species <20% relative abundance; (ii) contaminated *H. influenzae* or *H. parainfluenzae*, defined as one of these species at ≥50% relative abundance and a second species also highly represented (≥20% relative abundance); or (iii) other, where neither of these species exceeded 50% relative abundance (Fig. S1 online at doi.org/10.26188/14954931). Strain multiplicity for pure *H. influenzae* and *H. parainfluenzae* cultures was assessed by comparing the ratio of heterozygous to homozygous single nucleotide variant (SNV) calls (methods below) against an empirically determined threshold (*H. influenzae*, 0.025; or *H. parainfluenzae*, ≥0.100, calculated from public data sets); samples exceeding this threshold were considered mixed cultures and were excluded from further analysis (Fig. S1 online at doi.org/10.26188/14954931). Genomes were assembled with Unicycler v0.4.7 (77) using Illumina data in all cases and by MinION data where available. All AMR plasmid sequences (listed in Table S4 online at doi.org/10.26188/14954931) were identified as circularized contigs in the assembly graphs. Read data and assemblies were deposited under the NCBI BioProject accession PRJNA668428 (see Data Set S1 for individual accessions).

**Population structure analysis.** The *H. influenzae* and *H. parainfluenzae* short-read Illumina data generated in this study (*n* = 107), and publicly available read sets for previously sequenced genomes of these species (*n* = 891; summarized in Table S2 online at doi.org/10.26188/14954931), were subjected to SNV detection, phylogenetic, and population structure analyses. SNVs (biallelic and polyallelic) were called using Bowtie2 v2.2.9 (for read mapping) and SAMtools v1.9 (for variant calling) via the RedDog pipeline v1beta.11 (https://github.com/katholt/reddog), using *H. influenzae* strain Rd KW20 (accession GCA_00027305.1) and *H. parainfluenzae* strain T3T1 (accession GCA_000210895.1) as reference genomes. For each species, core SNV alleles were defined as SNV alleles present in ≥95% genomes. Maximum likelihood conserved core-genome SNV phylogenies were inferred from alignments of core SNV alleles (263,940 [85.3% of all detected SNV] for 901 *H. influenzae* genomes and 329,046 [79.1% of all detected SNV] for 97 *H. parainfluenzae* genomes) using IQ-TREE v2.1.0 (78). Phylogenies were visualized with ggtree v1.44.6 (79) in R v3.5.2 (80). *H. influenzae* capsular serotype loci were detected from genome assemblies using hicap v1.0.0 (42), and sequence types (STs) were assigned to *H. influenzae* read sets using SRST2 v2.0.2 (81) with the *H. influenzae* multilocus sequence typing (MLST) database (82) (https://pubmlst.org/organisms/haemophilus-influenzae/).

Discriminant analysis of principal components (DAPC) (83) was conducted to explore the relationship between bacterial population structure and sample source using k-mers of length *K* = 16 extracted from assemblies. Frequencies of k-mers were counted in each assembly with fm-lite v1.0 (https://github.com/nvalimak/fm-lite), and a presence-abscence matrix was constructed. Due to memory limitations, random sets of 500,000 k-mers were selected from the presence-absence matrix to use as input for DAPC, which was performed with the R package adegenet v2.1.1 (84) in triplicate using different random k-mer subsets to ensure stability of results (and additionally using the core-genome SNVs called from reads).

**Analysis of antimicrobial determinants.** RCH isolates were investigated for known and novel AMR determinants. Reads and assemblies were screened using SRST2 v2.0.2 and BLAST v2.7.1, respectively, to identify alleles of horizontally transferred AMR genes curated in the ARG-ANNOT database (85). Exact matches for translated blaTEM gene sequences were identified in the NCBI AMR database with BLAST to infer the spectrum of activity and inhibitor resistance.

Mutations in chromosomally encoded antimicrobial target genes (ftsI, folA, rpoB, and pbp genes) (43, 44, 46, 64, 86–92) were also investigated. An exhaustive search for PBP genes present in the *H. influenzae* and *H. parainfluenzae* reference genomes was performed by aligning translated gene sequences to all curated PBP protein sequences available in the Swiss-Prot database (93) to identify those with ≥80% coverage and ≥70% identity (Table S3 online at doi.org/10.26188/14954931). Nucleotide sequences for target genes (ftsI, folA, rpoB, and pbp genes) were extracted from RCH isolate assemblies, and the translated amino acid
sequences were aligned using MAFFT v7.407 (94). Each alignment position was compared to the consensus sequence for all isolates that were sensitive to the relevant antimicrobial. Positions that varied were tested for statistical association with the corresponding antimicrobial susceptibility phenotype (expressed as a binary variable, insensitive [I/R] versus sensitive [S]) using Fisher's exact test and using linear mixed models (LMMs) to correct for population structure by including a genetic relatedness matrix calculated from the distribution of variants in isolates are detailed in Data Set S2.

Identification of persistent and transmitted strains. Strains were defined as groups of closely related isolates with a pairwise genetic distance of ≥20. They were identified initially using complete-linkage hierarchical clustering based on SNV distances derived from the global conserved core-genome alignment (Fig. S2 online at doi.org/10.26188/14954931). To capture isolate pairs with inflated SNV counts due to small numbers of recombination events, the SNV distance thresholds for strain definition were set to ≥2,000 for *H. influenzae* and ≥4,000 for *H. parainfluenzae* (Fig. S2A and C online at doi.org/10.26188/14954931). Precise pairwise SNV distances within each potential strain group were obtained by mapping all isolates to the best within-group genome assembly (lowest *N*<sub>50</sub> value) using RedDog v1beta.11 as described above. Recombination blocks were identified by comparing pairwise SNV densities within discrete 4-kbp windows along the genome with the mean pairwise SNV count for all windows, using a binomial test and Bonferroni correction to account for multiple testing within each strain group (Fig. S3 online at doi.org/10.26188/14954931). Genetic distance between isolate pairs was then defined as nonrecombinant SNVs plus the number of recombinant blocks, and strains were defined as groups of isolates with pairwise genetic distance of ≥20 (Fig. S2B and D online at doi.org/10.26188/14954931).

Data availability. Read data and assemblies of *H. influenzae* and *H. parainfluenzae* isolates are available through NCBI BioProject under the accession PRJNA668428. Accessions for individual isolates are additionally listed in Data Set S1.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**DATA SET S1**, XLSX file, 0.1 MB.

**DATA SET S2**, XLSX file, 0.03 MB.

ACKNOWLEDGMENTS

This work was supported by the Bill & Melinda Gates Foundation, Seattle (K.E.H. [OPP1175797]), and the Australian Government Research Training Program (S.C.W.). K.E.H. was additionally listed in Data Set S1.

We declare that there are no conflicts of interest.

REFERENCES


