Virulence determinants of *Moraxella catarrhalis*: distribution and considerations for vaccine development

Luke V. Blakeway,1 Aimee Tan,1 Ian R. A. Peak1,2,* and Kate L. Seib1,*

Abstract

*Moraxella catarrhalis* is a human-restricted opportunistic bacterial pathogen of the respiratory mucosa. It frequently colonizes the nasopharynx asymptomatically, but is also an important causative agent of otitis media (OM) in children, and plays a significant role in acute exacerbations of chronic obstructive pulmonary disease (COPD) in adults. As the current treatment options for *M. catarrhalis* infection in OM and exacerbations of COPD are often ineffective, the development of an efficacious vaccine is warranted. However, no vaccine candidates for *M. catarrhalis* have progressed to clinical trials, and information regarding the distribution of *M. catarrhalis* virulence factors and vaccine candidates is inconsistent in the literature. It is largely unknown if virulence is associated with particular strains or subpopulations of *M. catarrhalis*, or if differences in clinical manifestation can be attributed to the heterogeneous expression of specific *M. catarrhalis* virulence factors in the circulating population. Further investigation of the distribution of *M. catarrhalis* virulence factors in the context of carriage and disease is required so that vaccine development may be targeted at relevant antigens that are conserved among disease-causing strains. The challenge of determining which of the proposed *M. catarrhalis* virulence factors are relevant to human disease is amplified by the lack of a standardized *M. catarrhalis* typing system to facilitate direct comparisons of worldwide isolates. Here we summarize and evaluate proposed relationships between *M. catarrhalis* subpopulations and specific virulence factors in the context of colonization and disease, as well as the current methods used to infer these associations.

INTRODUCTION

*Moraxella catarrhalis* is a human-restricted opportunistic bacterial pathogen of the respiratory tract. Colonization of the nasopharynx peaks in young children and steadily declines towards adulthood before rising again in the elderly [1]. Although it is able to colonize asymptomatically, *M. catarrhalis* is also an important aetiological agent of otitis media (OM) in infants and children, and acute exacerbations of chronic obstructive pulmonary disease (COPD) in the elderly and heavy smokers [2]. *M. catarrhalis* is the third most prevalent pathogen associated with OM worldwide, behind *Streptococcus pneumoniae* and nontypeable *Haemophilus influenzae* (NTHi) [3], and the second most frequently isolated pathogen associated with exacerbations of COPD, behind NTHi and equal with *S. pneumoniae* [4]. *M. catarrhalis* is also occasionally the causative agent of sinusitis, meningitis and conjunctivitis (reviewed in [5]). The progression from colonization to symptomatic disease does not occur in all cases of *M. catarrhalis* acquisition. A number of bacterial factors have been identified that may contribute to virulence, some of which are present in all isolates, while others are differentially distributed. Several of these factors have also been considered as potential vaccine candidates [6, 7]. It is unclear whether the expression of specific virulence factors contributes to a strain’s potential to cause disease, or to remain a commensal. Environmental and host factors also play an important role in the progression of *M. catarrhalis* disease [8]. The key function of *M. catarrhalis* virulence factors has recently been reviewed in detail [6, 7], and here we discuss relationships between *M. catarrhalis* subpopulations and specific virulence factors in the context of colonization and disease.

Received 2 June 2017; Accepted 14 August 2017

Author affiliations: 1 Institute for Glycomics, Griffith University, Gold Coast, Queensland, Australia; 2 School of Medical Science, Griffith University, Gold Coast, Queensland, Australia.

*Correspondence: Ian R. A. Peak, i.peak@griffith.edu.au; Kate L. Seib, k.seib@griffith.edu.au

Keywords: *Moraxella catarrhalis*; otitis media; chronic obstructive pulmonary disease; virulence factors; phase variation; vaccine.

Abbreviations: AFLP, single-adaptor amplified fragment length polymorphism; AOM, acute otitis media; COPD, chronic obstructive pulmonary disease; ECM, extracellular matrix; IL-8, interleukin 8; LOS, lipooligosaccharide; NTHI, nontypeable *Haemophilus influenzae*; OM, otitis media; OME, otitis media with effusion; OMVs, outer membrane vesicles; ORF, open reading frame; PCR-RFLP, polymerase chain reaction-restriction fragment length polymorphism; PCV, pneumococcal conjugate vaccine; PHID-CV, pneumococcal *Haemophilus influenzae* protein D conjugate vaccine; pRFLP, probe generated restriction fragment length polymorphism; RB, 16s ribotype; REA, restriction endonuclease analysis; RTI, respiratory tract infections; TNF-α, tumour necrosis factor α.
The importance of *M. catarrhalis* in otitis media

OM is a spectrum of related diseases that occur predominantly in children and is divided into two primary clinical presentations, acute otitis media (AOM) and otitis media with effusion (OME). AOM involves purulent inflammation of the middle ear with associated acute-onset local and systemic symptoms (otalgia, otorrhea and fever) [9], while OME involves the accumulation of middle ear fluid in the absence of acute symptoms of infection and often precedes AOM [10]. Poorly managed OM has severe consequences for afflicted children, including scarring of the tympanic membrane following repeated perforation, which can lead to acute or chronic hearing loss [11, 12] and subsequent difficulty in learning [13]. Worldwide studies have consistently revealed a high incidence of OM, with children under 5 in Oceania among the most affected [14]. OM is particularly prevalent in Indigenous Australian children, with one study reporting that as many as 95% of infants had already suffered AOM or OME in the first 8 weeks of life [15]. OM is the main cause of general practitioner consultations [16], antibiotic prescribing [17] and surgical procedures [18] in children in the developed world.

OM is typically a polymicrobial infection that predominantly involves at least one of three bacterial otopathogens, *S. pneumoniae, NTHi* and/or *M. catarrhalis*, and any of several upper respiratory tract viruses (e.g. rhinovirus, respiratory syncytial virus and influenza virus) (reviewed in [19]). Recent studies have shown that the prevalence of *M. catarrhalis* in OM is as high as 20% by culture [20], and this rate increases substantially with the use of PCR, suggesting that *M. catarrhalis* is much more prevalent in OM than originally thought [20, 21]. *M. catarrhalis* is often more frequently detected in ear fluids as part of a co-infection with *S. pneumoniae* or *H. influenzae* than alone [21, 22]. The concurrent nasopharyngeal colonization of *M. catarrhalis* with *S. pneumoniae* or *H. influenzae* increases the risk of OM development compared to infection with any pathogen alone [23]. The relationship between *M. catarrhalis* co-infection and increased incidence of OM has also been observed in mouse [24] and chinchilla models [25]. The reason for this increase in incidence is not completely understood, however synergistic interactions between *M. catarrhalis* and other middle-ear pathogens have been demonstrated experimentally. *M. catarrhalis* is present in polymicrobial biofilms in the middle ear [26], and biofilm formation has been linked to resistance to treatment [25]. *M. catarrhalis* protects susceptible *S. pneumoniae* [25, 27] and *H. influenzae* strains [28] from β-lactam antibiotics in polymicrobial biofilms, likely mediated by *M. catarrhalis* β-lactamase containing outer-membrane vesicles (OMVs) [29]. In addition, *M. catarrhalis* OMVs were also shown to protect susceptible *H. influenzae* strains from complement-mediated killing [30]. Similarly, *S. pneumoniae* increases *M. catarrhalis* resistance to azithromycin in biofilm [25], while *H. influenzae* promotes *M. catarrhalis* persistence in biofilms in a quorum sensing-dependent manner [28].

The importance of *M. catarrhalis* in COPD

COPD is a progressive airway disease consisting of persistent airflow limitation and chronic inflammation [31], and is currently the fourth greatest cause of death worldwide [32]. COPD sufferers experience alternating periods of stability and acute exacerbations of the disease characterized by the sudden worsening of respiratory symptoms (sputum production, purulence and tenacity, cough and dyspnea) [33]. Frequent exacerbations have a detrimental impact on lung function [34] and quality of life [35], and drastically increase the risk of mortality [36]. *M. catarrhalis* accounts for approximately 10% of all exacerbations of COPD and causes approximately 2–4 million exacerbations per year in the USA [37].

The role of *M. catarrhalis* in the progression of the disease is not completely understood. Similar to OM, *M. catarrhalis* is frequently isolated from sputum samples of COPD patients in co-infections with *S. pneumoniae* and *H. influenzae* [38], although relationships between the incidence of exacerbation and co-infection have not been described. Although causality has not been confirmed, the onset of exacerbation is associated with the acquisition of new strains of *S. pneumoniae, H. influenzae* and *M. catarrhalis* [39]. A prospective longitudinal study showed that approximately 50% of acquisitions of a new strain of *M. catarrhalis* are associated with exacerbation [37]. Exacerbations coinciding with the acquisition of a new strain of *S. pneumoniae, H. influenzae* or *M. catarrhalis* are associated with an increased airway and systemic inflammatory response (including increased levels of IL-8, neutrophil elastase, TNF-α and C-reactive protein) compared to baseline levels, or exacerbations occurring with pre-existing strains, atypical bacteria or no bacterial presence [40]. This imbalance of proteases (e.g. neutrophil elastase) and anti-proteases (e.g. secretory leukocyte protease inhibitor) is hypothesized to be the major cause of lung tissue damage in COPD [41].

One study compared asymptomatic colonizing strains of *H. influenzae* with strains isolated from periods of exacerbation. The exacerbation-associated strains induced higher production of the inflammatory cytokine IL-8 by primary human tracheobronchial epithelial cells in vitro and also increased neutrophil recruitment in the lungs of mice [42]. This supported a previous suggestion that the likelihood of an exacerbation occurring upon acquisition of a new strain is related to the virulence of an acquired strain, with more virulent strains causing a greater immune response than colonizing strains [39]. However, the differences in the levels of inflammatory markers caused by infection with *M. catarrhalis* strains isolated from colonization versus those isolated from exacerbations of COPD have not been studied.

Treatment options

Current treatment options for OM and exacerbations of COPD rely on antibiotics, but the development of efficacious vaccines against *M. catarrhalis*, NTHi and *S. pneumoniae* is critical to reducing the burden of these diseases.
OM, antibiotics are only modestly more effective than no treatment, but cause adverse effects in 4–10 % of children [43]. The major OM pathogens are capable of forming biofilm in the middle ear [44] and orally administered antibiotics do not reach the predicted levels required to eradicate middle-ear biofilms [45]. This may explain why approximately 25 % of children with AOM do not respond to antibiotic therapy [46]. The benefit of antibiotics in the treatment of exacerbations of COPD is inconsistent outside of very severe cases [47], and antibiotic use is associated with an increased risk of harbouring antibiotic-resistant organisms in COPD patients [48]. *M. catarrhalis* remains susceptible to the majority of antimicrobials recommended for the treatment of OM and exacerbations of COPD [49, 50], although the near ubiquitous acquisition of β-lactamase by *M. catarrhalis* has rendered >90 % of isolates resistant to penicillins (reviewed in [5]).

The widespread use of the 7-, 10- and 13-valent pneumococcal conjugate vaccines (PCVs) in the USA, Europe, the UK and Australia has been associated with a decline in AOM of approximately 20 % [51, 52], but it has also resulted in a significant change in the microbiology of OM, with increased proportions of OM being associated with non-vaccine *S. pneumoniae* serotypes, as well as NTHi and *M. catarrhalis* [53]. Whilst no vaccine is licensed for NTHi, a licensed pneumococcal vaccine containing NTHi protein D conjugated to a 10-valent pneumococcal vaccine is available (PHID-CV, Synflorix) and an NTHi vaccine for COPD has recently been tested in clinical trials [54]. In contrast, there are currently no vaccines licensed that specifically target *M. catarrhalis*. Numerous vaccine candidates for *M. catarrhalis* have been proposed, but none have progressed to clinical trial and no correlates of protection have been identified [6]. Therefore, it is not known what type of immune response is required to protect against disease, and whether vaccine-induced antibodies need to be bactericidal, or if antibody blocking of antigen function may be necessary. For example, bacterial attachment to host cell receptors and colonization of the mucosal surface is the first step of infection, and blocking these interactions with vaccine-induced antibodies may be an effective strategy to prevent disease caused by *M. catarrhalis*. Further investigation of the distribution of virulence factors in *M. catarrhalis* is required so that vaccine development may be targeted at antigens that are conserved among disease-causing strains.

**DETERMINANTS OF *M. CATARRHALIS* VIRULENCE**

The question of whether different strains or subpopulations of *M. catarrhalis* vary in their propensity to cause disease, or if particular strains are better adapted to infect different host microenvironments (i.e. the nasopharynx, middle ear or lung) and cause OM as opposed to exacerbations of COPD or vice versa, is open to debate. *M. catarrhalis* isolates vary with respect to a range of *in vitro* phenotypes: e.g. adherence to epithelial cells [55, 56], resistance to killing by human serum [57], or antibiotic resistance [56, 58]. However, clear relationships between *in vitro* *M. catarrhalis* phenotypes and clinical presentation have not been established. Several distribution studies have investigated the presence or absence of specific genes and the proteins they encode in panels of *M. catarrhalis* isolates that are representative of different age groups, genetic lineages and carriage versus disease statuses (see Tables 1 and 2). However, no single virulence determinant or phenotype has been associated with either disease or carriage status. The challenge of determining which of the proposed *M. catarrhalis* virulence factors are relevant to disease is amplified by the lack of a widely used, standardized *M. catarrhalis* typing system that facilitates direct comparisons of worldwide isolates and associated epidemiological information (the typing systems are summarized in Box 1).

**Genetic lineage and virulence**

The *M. catarrhalis* species consists of two distinct phylogenetic lineages. The major lineage, 16 s ribotype 1 (RB1), includes 80–92 % of isolates, while the remaining 7–20 % of isolates form the minor RB2/3 lineage, split between 16 s ribotype 2 and 16 s ribotype 3 [55, 59–61]. The RB1 lineage was associated with virulence by Wirth *et al.*, who found that 51 % of RB1 lineage isolates were disease-associated, compared to only 14 % of RB2/3 lineage isolates [62]. However, in a similar study by Bootsma *et al.*, the RB1 and RB2/3 lineages were distributed evenly across carriage and disease isolates [55]. Resistance to complement-mediated killing is assumed to be an important virulence trait of *M. catarrhalis*, and both Wirth *et al.* and Bootsma *et al.* found that serum-resistant isolates, for which the ribotype was determined, were more frequently of the RB1 lineage [55, 62]. The separation of serum-resistant and serum-sensitive strains into distinct genetic lineages was also found by Verduin *et al.* [59], however, a clear relationship between serum resistance, genetic lineage and disease is not clear across studies. Hol *et al.* found that child-carriage isolates were serum-resistant less frequently than adult disease isolates (41.5 vs 89 %) [57], but this association was not observed by others, who found that serum resistance is consistently high (83–95 %) in *M. catarrhalis* isolates regardless of age group or clinical presentation [60, 63]. Recent supragenome modelling of 31 *M. catarrhalis* strains by Earl *et al.* demonstrated that the majority of well-characterized *M. catarrhalis* virulence factors were present in the core genome of serum-sensitive and serum-resistant strains belonging to either the RB1 or RB2/3 lineages, suggesting that all strains of *M. catarrhalis* are equally able to cause disease [64].

**M. catarrhalis virulence factors and potential vaccine candidates**

**UspA1**

UspA1 is an outer-membrane protein that is involved in adherence to epithelial cells [65, 66], the extracellular matrix (ECM) [67, 68] and biofilm formation [69]. Studies have indicated a significant association of uspA1 with the RB1
Table 1. Virulence factors and potential vaccine candidates with heterogeneous distribution or variable sequence

<table>
<thead>
<tr>
<th>Virulence factor/vaccine candidate</th>
<th>Function</th>
<th>Gene presence</th>
<th>Conservation identity†</th>
<th>Expression</th>
<th>Association†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>%</td>
<td>(n)*</td>
<td>%</td>
<td>(n)*</td>
</tr>
<tr>
<td>Ubiquitous surface protein A1</td>
<td>Adherence to epithelial cells [65, 66]</td>
<td>99%</td>
<td>(193/195) [60]</td>
<td>Modular [70, 71]</td>
<td>97%</td>
</tr>
<tr>
<td></td>
<td>Asdherence to ECM [67, 68]</td>
<td>97%</td>
<td>(109/112) [61]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Biofilm formation [69]</td>
<td>99%</td>
<td>(107/108) [63]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ubiquitous surface proteins A2/2H/2V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UspA2 allele</td>
<td>Serum resistance [66, 75]</td>
<td>72%</td>
<td>(141/195) [60]</td>
<td>Modular [70]</td>
<td>72%</td>
</tr>
<tr>
<td></td>
<td>Adherence to ECM [67, 68]</td>
<td>76%</td>
<td>(85/112) [61]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>77%</td>
<td>(83/108) [63]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UspA2H allele</td>
<td>Adherence to epithelial cells [65]</td>
<td>20%</td>
<td>(39/195) [60]</td>
<td>Modular [70]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Biofilm formation [76]</td>
<td>15%</td>
<td>(17/112) [61]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Serum resistance [75]</td>
<td>21%</td>
<td>(23/108) [63]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UspA2Y allele</td>
<td>Adherence to epithelial cells [77]</td>
<td></td>
<td></td>
<td>Modular [77]</td>
<td>10%</td>
</tr>
<tr>
<td>MID/Hag</td>
<td>Haemagglutination [80]</td>
<td>90%</td>
<td>(176/195) [60]</td>
<td>65.3%</td>
<td>(5) [83]</td>
</tr>
</tbody>
</table>
|                                   | Non-immune IgD binding [80] | 80% | (90/112) [61] | 56.6% | (7) [134] | 83% | (10/12) [80] | ✓ | ✓ | /
|                                   | Adherence to epithelial cells [81, 82] | 100% | (98/98) [83] | 60% | (4) [80] | | | | |
| CopB                              | Iron acquisition [84] | 100% | (90/90) [55] | 76.1% | (14) [87] | 100% | (24/24) [87] | ✓ | × | ? |
|                                   | Serum resistance [85] | 100% | (195/195) [60] | | | | | | |
| Opa-like protein                   | Unknown [128] | 100% | (27/27) [128] | 43% | (7) [128] | 100% | (25/25) [128] | ✓ | ✓ | /
| Moraxella haemagglutinin-like proteins | Adherence to epithelial cells [132] | 100% | (20/20) [133] | 90.1% | (3) [132] | 63% | (10/16) [132] | ✓ | ✓ | /C2 |
|                                   | Adherence to epithelial cells [132] | 100% | (20/20) [133] | 68.8% | (2) [132] | 63% | (10/16) [132] | ✓ | ✓ | /
|                                   | Adherence to epithelial cells [132] | 100% | (20/20) [133] | 98% Nt | (7) [132] | 63% | (10/16) [132] | ✓ | ✓ | /
|                                   | Lactoferrin-binding protein B | 100% | (3/3) [125] | 77% | (3) [125] | 100% | (90/90) [125] | ✓ | ✓ | /
|                                   | Transferin-binding protein B | 100% | (6/6) [120] | 51% | (6) [120] | 100% | (90/90) [120] | ✓ | ✓ | /
|                                   | Type IV pilin (PHA subunit) | 100% | (106/106) [126] | 59% | (22) [126] | 100% | (88) [126] | ✓ | ✓ | /
|                                   | DNA methylation [110] | 100% | (81/81) [110] | 70% | (3) [110] | | | | |
|                                   | β-lactamase | 95% | (1377/1440) [99] | 99.7% | (2) [103] | 92% | (385/419) [104] | ✓ | ✓ | /
|                                   | Lactoferrin-binding protein B | 98% | (113/115) [86] | 91% | (375/413) [102] | ✓ | ✓ | /
|                                   | Lipooligosaccharide (LOS) | 94% | (184/195) [60] | | | | | | |
|                                   | Invasion of epithelial cells [106] | 96% | (110/115) [86] | | | | | | |

*Number of isolates examined in each study.
†Minimum % amino acid identity unless otherwise indicated (Nt, nucleotide). uspA genes are highly variable in sequence due to their modular arrangement. See text for details.
‡Lin, genetic lineage (RB1 or RB2/3); Age, age group (child or adult); Clin, clinical manifestation (carriage, OM, or COPD); ✓, association observed; ×, no association observed; ?, debated association; blank, no information available. See text for details.
<table>
<thead>
<tr>
<th>Virulence factor/vaccine candidate</th>
<th>Function</th>
<th>Gene presence</th>
<th>Conservation identity†</th>
<th>Expression</th>
<th>Association‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outer-membrane protein CD</td>
<td>Adherence to epithelial cells [88, 89] Adherence to mucus [90] Membrane stability [89]</td>
<td>100% (195/195) [60] (30/30) [91] (98) [92]</td>
<td>97.1% (9) [92]</td>
<td>100% (51/51) [93]</td>
<td>✓ x ?</td>
</tr>
<tr>
<td>Outer-membrane protein E</td>
<td>Putative nutrient acquisition [91] Serum resistance [91]</td>
<td>70% (81/115) [86]</td>
<td>96.6% (16) [96]</td>
<td>100% (19/19) [97]</td>
<td>x ?</td>
</tr>
<tr>
<td>Outer-membrane proteins G1a and G1b</td>
<td>Putative copper transport [115]</td>
<td>100% (25/25) [115]</td>
<td>90% (25) [115]</td>
<td>x x</td>
<td></td>
</tr>
<tr>
<td>OmpG1a</td>
<td>Unknown [115]</td>
<td>100% (25/25) [115]</td>
<td>92% (25) [115]</td>
<td>x x</td>
<td></td>
</tr>
<tr>
<td>OmpG1b</td>
<td>Unknown [98]</td>
<td>100% (12/12) [61] (96/96) [98]</td>
<td>90% Nt (14) [98]</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>M35</td>
<td>100% (25/25) [115]</td>
<td>92% (25) [115]</td>
<td>100% (14/14) [116]</td>
<td>x x</td>
<td></td>
</tr>
<tr>
<td>Oligopeptide permease A</td>
<td>Peptide binding [139]</td>
<td>100% (21/21) [127]</td>
<td>98.7% Nt (21) [127]</td>
<td>x x</td>
<td></td>
</tr>
<tr>
<td>Moraxella surface proteins 22, 75 and 78</td>
<td>Heme binding [140]</td>
<td>100% (25/25) [117]</td>
<td>99% (10) [117]</td>
<td>100% (8/8) [141]</td>
<td>x x</td>
</tr>
<tr>
<td>Msp22</td>
<td>Putative succinic dehydrogenase [117]</td>
<td>100% (25/25) [117]</td>
<td>97% (10) [117]</td>
<td>100% (8/8) [141]</td>
<td>x x</td>
</tr>
<tr>
<td>Msp75</td>
<td>Putative nitrate reductase [117]</td>
<td>100% (25/25) [117]</td>
<td>99% (10) [117]</td>
<td>x x</td>
<td></td>
</tr>
<tr>
<td>Substrate-binding protein 2</td>
<td>Arginine uptake [142]</td>
<td>100% (30/30) [118] 99.8% Nt (30) [118]</td>
<td>100% (11/11) [118]</td>
<td>x x</td>
<td></td>
</tr>
<tr>
<td>CySP</td>
<td>Sulfate and thiosulfate ion binding [121]</td>
<td>100% (18/18) [121]</td>
<td>96% (18) [121]</td>
<td>x x</td>
<td></td>
</tr>
<tr>
<td>AfeA</td>
<td>Manganese, zinc, ferrous and ferric ion binding [122]</td>
<td>100% (66/66) [122]</td>
<td>87% (46) [122]</td>
<td>100% (20/20) [122]</td>
<td>x x x</td>
</tr>
<tr>
<td>Lactoferrin-binding protein A</td>
<td>Lactoferrin utilization [119]</td>
<td>100% (2/2) [125]</td>
<td>99% (2) [133]</td>
<td>100% (8/8) [125]</td>
<td>x x</td>
</tr>
<tr>
<td>Transferrin-binding protein A</td>
<td>Putative transferrin binding [120]</td>
<td>100% (2/2) [120]</td>
<td>98% (2) [120]</td>
<td>100% (32/32) [120]</td>
<td>x x</td>
</tr>
</tbody>
</table>

*Number of isolates examined in each study.
†Minimum % amino acid identity unless otherwise indicated (Nt, nucleotide).
‡Lin, genetic lineage (RB1 or RB2/3); Age, age group (child or adult); Clin, clinical manifestation (carriage, OM, or COPD); ✓, association observed; x, no association observed; ?, debated association; blank, no information available. See text for details.
BOX 1. M. catarrhalis typing systems

Serotyping

M. catarrhalis is acapsular and is the only serotyping system developed for the species to date that differentiates isolates by the structure of outer-membrane lipooligosaccharide (LOS). In contrast to the diversity of LOS found in other respiratory pathogens (e.g. NTHi [143]), the M. catarrhalis LOS is much less heterogeneous (serotype A approximately 60–75 %, B approximately 20–30 %, C approximately 2–6 %, with approximately 5 % untypeable) [60, 86, 107, 108]. Thus, LOS serotyping is infrequently used for typing M. catarrhalis.

Proteomic categorisation

Many early efforts to differentiate M. catarrhalis were protein-based methods using variations of whole-cell protein electrophoresis and observed varying degrees of heterogeneity in protein expression between strains [144–147]. Small studies focusing on differentiating strains based on biochemical testing [148] or select proteins (e.g. β-lactamases [149]) were also performed, although these methods were supplanted by genetic typing systems. Differentiation of M. catarrhalis based on outer-membrane proteins has been revisited more recently using 2D protein analysis and MALDI-TOF MS [150] and has highlighted diversity between the genetic lineages of M. catarrhalis.

Genetic methods

The major genetic division of M. catarrhalis strains is based on the 16 s ribotyping method, which divides strains into three ribotypes (RB1, RB2 and RB3) divided into two lineages (RB1 and RB2/3) based on the sequence of the 16 s rRNA gene [55, 59–61]. The diversity of M. catarrhalis has been further elucidated by more recent molecular genotyping techniques with greater discriminatory power and reproducibility, however a lack of standardization between worldwide laboratories has made direct comparisons between studies difficult. M. catarrhalis typing methods based on agarose gel electrophoresis of fragmented chromosomal DNA (e.g. restriction endonuclease analysis (REA) [151] and pulsed-field gel electrophoresis (PFGE) [152]) are the most commonly used, with DNA hybridization (e.g. multi-locus Southern blotting [153] and probe-generated restriction fragment length polymorphism (pRFLP) [55]) and PCR-based methods (random amplified polymorphic DNA (RAPD) [154], single-adaptor amplified fragment length polymorphism (sAFLP) [55] and PCR-restriction fragment length polymorphism (PCR-RFLP) [153]) also used to a limited degree. An MLST scheme has been developed for M. catarrhalis [62] (found at https://enterobase.warwick.ac.uk/) and the associated database currently contains information for 555 strains. Routine use of MLST for M. catarrhalis has the potential to allow for the investigation of relationships between sequence type and clinical source, although its current usefulness is limited by incomplete epidemiological information for the submitted strains. More recently, whole-genome sequencing has been used with 35 strains to show population structure [64, 155], but current costs make this impractical for use as a standard typing system.

Collectively, studies utilizing genetic typing methods could not detect any significant geographical clustering of genotypes [62, 156] and no relationship of a particular genotype to disease has been noted; isolates from identical clinical presentations frequently display diverse genotypes, while isolates indistinguishable by PCR-RFLP and PFGE have also been isolated from different disease presentations [153]. In a longitudinal study of nosocomial infection, no genotypes with an increased propensity for transmission were identified [157].

Summary of typing information

Numerous typing techniques have been used to differentiate M. catarrhalis isolates, and have collectively shown that the species is genetically diverse, with a large number of distinct strains in circulation [156, 158]. No consensus has yet been reached on an optimal typing system for the species, although there is increasing need for standardization so that the results of worldwide studies may be compared directly. Sequencing-based typing systems (e.g. MLST and/or multi-virulence locus sequence typing) are currently the most promising due to the discrete nature of the data generated and should be used more frequently, provided that they are accompanied by an accessible database that permits the inclusion of associated epidemiological data so that relationships may be inferred between virulence factors/ vaccine candidates and disease.

Lineage: uspA1 is present in 87–98 % of RB1 isolates and only 23–36 % of RB2/3 lineage isolates [55, 62]. However, studies by Verhaegh et al. showed that the presence of uspA1 was high in both the RB1 (100 %) and RB2/3 (>89 %) lineages, regardless of clinical manifestation or age group [60, 61]. The uspA1 gene is modular and the CEACAM- and fibronectin-binding motifs, which confer UspA1-mediated adherence, are not conserved in all UspA1-expressing M. catarrhalis isolates [70, 71]. Adding to the complexity, uspA1 is phase-variable, with this mediated by a homopolymeric G tract upstream of its open reading frame (ORF), changes in which alter the levels of expression of
UspA1 protein and the level of adherence to conjunctival epithelial cells in vitro [72]. These variations in sequence and expression levels may affect whether uspA1 is associated with clinical presentation of M. catarrhalis, however this is not well investigated. For example, while UspA1 was shown to be expressed in 95% of child-carriage isolates [63], 97% of child respiratory tract infection (RTI) and 95% of adult RTI isolates [60], neither sequence variation nor levels of UspA1 expression (high or low) were described in either of these studies. One study showed that child-carriage isolates in which UspA1 expression was not detected belong to the RB2/3 lineage [73], but a later study by the authors showed that detectable levels of UspA1 expression in these isolates could be induced by cold shock at 26°C (a temperature commonly experienced by the human nasopharynx), compared to growth at 37°C [74]. Consequently, any association with uspA1 variation or expression with age group and clinical manifestation warrants further investigation.

**UspA2, UspA2H and UspA2V**

The uspA2 locus can contain one of three mutually exclusive alleles known as uspA2, uspA2H and uspA2V, each encoding a different outer-membrane protein with functional and distributional differences and limited conservation. uspA2 is found in approximately 72–77% of isolates [60, 61, 63] and is involved in serum resistance [66, 75] and adherence to the ECM [67, 68]. uspA2H is found in approximately 15–21% of isolates [60, 61, 63], and is considered a hybrid of uspA1 and uspA2, with a role in adherence [65], biofilm formation [76] and serum resistance in vitro [75]. uspA2V is found in approximately 10% of isolates and is also potentially involved in adherence and serum resistance [77]. Despite the importance of uspA2 in serum resistance, uspA2 is equally present in both serum-resistant and serum-sensitive isolates and is not associated with either ribosomal lineage [55]. All uspA2H PCR-positive isolates belong to the RB1 lineage, irrespective of age group and specimen source [60], while uspA2V is associated with the RB2/3 lineage [77]. However, until more information is available on the function of UspA2V, the relevance of this association for the capacity of strains to cause disease is unclear. A significant difference in the distribution of uspA2 and uspA2H was observed in RTI isolates from children (95% uspA2, 5% uspA2H) compared to adults (61% uspA2, 39% uspA2H) [60]. A difference, albeit smaller, is also seen when comparing the uspA2/2H distribution in child or adult RTI isolates above to the distribution in child-carriage isolates (77–83% uspA2 and 17–21% uspA2H) in studies performed by the same [61] and other authors [63]. Taken together, this may suggest a role for UspA2 in childhood disease or for UspA2H in the infection of individuals who have developed an immune response to UspA2. Similar to uspA1, uspA2, uspA2H and uspA2V are modular, resulting in sequence variations within a particular allele that potentially alters function [70, 77]. For example, serum resistance is not conferred by all UspA2 proteins [70].

In addition to sequence differences, uspA2 and uspA2H exhibit phase-variable expression. A 5′-(AGAT)n-3′ tetranucleotide repeat tract is present upstream of the ORF of uspA2, and experimental alterations of the uspA2 repeat tract have shown that maximal serum resistance is conferred when >9 repeat units are present, while mutants with <9 repeats are serum-sensitive [78]. Consistent with this, the expression of UspA2 in clinical isolates is variable and correlates with serum resistance through C3 binding [79]. A polyA tract is present in the uspA2H ORF, and on-off switching of the expression of UspA2H alters the ability of M. catarrhalis to auto-aggregate, resist complement-mediated killing and adhere to epithelial cells [60]. Little is known about the expression of UspA2V [77]. Further investigation of the sequence variation and expression of UspA2, UspA2H and UspA2V in relation to clinical manifestation in longitudinal studies is required to determine the relevance of these proteins in disease and the conservation of expression for vaccine development purposes.

**MID/Hag**

MID/Hag is located in the outer membrane and mediates haemagglutination and non-immune binding of IgD by M. catarrhalis [80], while it also functions as an adhesin for cells derived from the human lung, middle ear and ciliated bronchial epithelium [81, 82]. mid/hag was found in 80% of child-carriage isolates [61], 90% of child RTI and 91% of adult RTI isolates [60]. Another study showed that mid/hag was present in 100% of isolates from various clinical presentations [83], indicating that there was no association between gene presence and carriage or disease. However, also like the UspA proteins, MID/Hag is subject to phase variation, mediated by a polyG tract in its ORF, resulting in the presence or absence of functional MID/Hag and the respective adherence phenotype [83]. In isolates that have a mid/hag gene, MID/Hag is more frequently expressed by child RTI isolates than adult RTI isolates (92 vs 73%) [60]. It is unknown whether the association of MID/Hag expression with child versus adult disease isolates is due to phase variation, and further investigation is required. This may suggest that while MID/Hag expression bestows a selective advantage during infection of the child host, selection against MID/Hag expressing isolates may occur in adults following an adaptive immune response, with the phase variation of MID/Hag contributing to immune evasion. The presence of the mid/hag gene was also associated with the RB1 lineage in isolates from child carriage [61], and isolates from children and adults presenting with respiratory disease [60].

**CopB**

CopB is involved in iron acquisition [84] and serum resistance [85]. Most studies indicate that all isolates possess the copB gene [55, 60], although one study showed that copB is more frequently present in M. catarrhalis isolated from children and adults with RTI (50%) than in carriage isolates from children (0%) [86]. The reason for this difference in distribution is not clear, but may reflect localized...
geographical variance of *M. catarrhalis*. PCR restriction fragment length polymorphism analysis of the *copB* gene showed that allelic variation occurs between isolates, with five *copB* types described (*copB-I* and *copB-III*, 51%; *copB-II*, 38%; *copB-IV*, 9%; *copB-O*, 2%) [87]. A significant association was observed between *copB* types and ribosomal lineage. Isolates containing *copB* types I/III and II are associated with the RB1 lineage (99%), and *copB* types 0 and IV are associated with the RB2/3 lineage (86%) [60]. It is unknown whether allelic differences in *copB* have functional relevance or contribute to any differences in the capacity of isolates to cause disease, however antibodies directed at major surface epitopes do not bind all *CopB* types and this may have implications if CopB is pursued as a vaccine antigen [87].

**Outer-membrane protein CD (OmpCD)**

OmpCD is an outer-membrane protein that is involved in adherence to epithelial cells [88, 89], adherence to human nasopharyngeal, tracheobronchial and middle-ear mucins [90], and membrane stability [89]. One study showed that *ompCD* is more frequently present in *M. catarrhalis* isolated from children and adults with RTI (84%) than in carriage isolates from children (14%) [86]. However, other studies have shown that *ompCD* is present [60, 91, 92] and expressed [93] in all *M. catarrhalis* isolates from diverse clinical and geographical sources tested to date. Although the amino acid sequence of OmpCD is well conserved [92], PCR-RFLP analysis of the *ompCD* gene has grouped variants into two [60] or five [92] types, depending on the restriction endonuclease used. The two types described by Verhaegh et al. ([*ompCD1*, 91%; *ompCD2*, 9%] show variation within the first A549 epithelial cell-binding domain of the *ompCD* gene between isolates [60], however the functional relevance of this variation has not been investigated. In addition, a significant association was found between *ompCD1* and the RB1 lineage, while *ompCD2* is associated with the RB2/3 lineage [60].

**Outer-membrane protein J (OmpJ)**

OmpJ is an outer-membrane protein of unknown function [98]. Two *ompJ* types have been identified (*ompJ1*, 17%; *ompJ2*, 83%) that are >90% identical in nucleotide sequence [98]. *ompJ1* is associated with RB2/3 lineage isolates and *ompJ2* is associated with the RB1 lineage [61]. In addition, *ompJ1* was found primarily in serum-sensitive isolates, while *ompJ2* was almost exclusively associated with serum-resistant isolates [98]. However, serum survival assays determined that it was unlikely that OmpJ is directly involved in serum resistance [98].

**BRO β-lactamases**

*M. catarrhalis* β-lactamase mediates resistance to β-lactam antibiotics (e.g. penicillin and amoxicillin) and is encoded by two alleles, *bro1* and *bro2*, that are present in 82–91% and 4–11% of worldwide isolates, respectively [86, 99, 100]. A higher frequency of β-lactamase-positive isolates was observed in children compared to adults, although no difference was seen in the distribution of *bro* alleles between these groups [99]. However, *bro1* was found more frequently in disease isolates from adults and children compared to child carriage isolates (93–97% versus 75–82%), whereas the *bro2* gene was found more frequently in healthy carriers (14–16%) than disease isolates (5–6%) [86, 100]. All BRO-2 isolates were associated with the RB1 lineage [55]. The biological relevance of this is unclear, as although the minimal inhibitory concentration against β-lactam antibiotics is generally reported to be higher for BRO-1 than BRO-2 isolates, large MIC ranges have been observed for both, which often overlap [100–102]. The *bro* alleles differ by one amino acid, which has unknown significance, although *bro2* has a 21 base pair deletion in its promoter region that results in decreased expression compared to *bro1* and may account for the observed differences in MIC [103, 104].

**Lipoooligosaccharide (LOS)**

*M. catarrhalis* LOS is a surface-exposed glycolipid found in the outer membrane and is involved in adherence to [105] and invasion of epithelial cells [106], and serum resistance [105]. *M. catarrhalis* expresses three mutually exclusive LOS serotypes with varying distribution in the population: 60–75% of isolates are serotype A, 20–30% are serotype B and 2–6% are serotype C, while 5% of isolates are untypeable [60, 86, 107, 108]. Although no difference is observed in the distribution of LOS serotypes between carriage and disease isolates [86], a higher frequency of LOS type B and a lower frequency of LOS type A were found in respiratory isolates from adults compared to children [60]. In addition, all LOS type B isolates are within the RB1 lineage [60].

**ModM**

ModM is a phase-variable type III DNA methyltransferase that acts as an epigenetic regulator. ModM phase variation is mediated by a 5’-(CAAC)₃-3’ repeat tract in its ORF [109], and switching of ModM expression alters the expression of a distinct set of genes, known as a phase-variable regulon or phasevarion, via differential methylation of the
Other virulence factors and vaccine candidates

The literature suggests that there are a number of other virulence factors and vaccine candidates that are found in all *M. catarrhalis* isolates (Table 2). However, much of the evidence for the conservation and expression of these factors (e.g. outer-membrane proteins G1a and G1b [115], *M. catarrhalis* adherence protein [116], *Moraxella* surface proteins 22, 75 and 78 [117], substrate-binding protein 2 [118], lactoferrin-binding protein A [119] and transferrin-binding protein A [120]) is derived from studies that only examined a small number of isolates, or in the case of CysP [121] and AfeA [122], what can be gleaned from genomes available online. Consequently, further investigation in larger panels that include isolates from varying genetic lineages, age groups or clinical manifestations is needed to fully elucidate the suitability of these potential vaccine candidates. For example, M35 was considered to be highly conserved when sequences from RB1 isolates were compared [123], but sequence variations became apparent when M35 genes from RB2/3 isolates were sequenced [124].

For other virulence factors and potential vaccine candidates that are found ubiquitously in tested isolates, significant sequence variability is present, and whether these differences are correlated with lineage, age group, or clinical manifestations requires additional investigation. For example, lactoferrin-binding protein B from isolates of diverse geographic regions and clinical presentations shared 77% identity [125], and only 51% identity was observed in transferrin-binding protein B [120]. There are several type IV pilin PilA subunit alleles in circulation with as little as 59% amino acid sequence identity that are grouped into three clades (pilA1a, 32%; pilA1b, 10%; pilA2, 58%) [126]. Pilin clades are not associated with age group or disease, being evenly distributed between OM isolates from children and adult sputum isolates. However, all isolates examined for pilA type are of the RB1 lineage, and whether this distribution differs to that found in RB2/3 isolates has not been investigated [127]. For oligopeptide permease A (OppA), single-nucleotide polymorphisms at specific sites are found exclusively among OM isolates, while distinct set of mutations occur exclusively in sputum isolates [127], however the functional relevance of these polymorphisms is unknown. Two Opa-like protein A (olpA) alleles were identified in a small panel of isolates (olpA1, 93%; olpA2, 7%) with 43% identity between variants, and investigation of the distribution of these alleles in a larger study is warranted [128].

Data on other potential *M. catarrhalis* virulence factors and vaccine antigens, such as ORF113 [129], haemin-utilization protein (HumA) [130] and *Moraxella* haemoglobin-utilization protein (MhuA) [131] are currently lacking, and the distribution and conservation of these genes needs to be further examined before vaccine development can proceed. Although they are considered to be conserved, the gene presence and expression of *Moraxella* haemagglutinin-like proteins A1 and A2 was detected by targeting sequences present in both genes/proteins and their true distribution is yet to be determined [132, 133].

CONCLUSIONS AND FUTURE PERSPECTIVES

*M. catarrhalis* is a prevalent colonizer of the human respiratory tract and an important causative agent of OM in children and exacerbations of COPD in the adult population. The development of a *M. catarrhalis* vaccine is warranted to decrease the burden of disease caused by this pathogen. Many vaccine targets have been proposed for *M. catarrhalis*, but no consensus has been reached on which to move forward with. Crucially for vaccine development, no correlates of protection have been described. Whilst *M. catarrhalis* does not thus far appear to have the same level of antigenic variability as the other bacterial otopathogens, *S. pneumoniae* and NTHi, variability in the distribution and sequence of virulence determinants and vaccine candidates exists (Table 1), and in many cases has yet to be fully explored (Table 2). A cocktail of antigens to cover as broad a range of strains as possible is one potential solution, but another option may be to identify and target key virulence factors involved in colonization or disease. In either case, a clear understanding of the links between *M. catarrhalis* virulence determinants and disease is required.

Links between sub-populations of *M. catarrhalis* and specific virulence factors or the capacity to cause disease have been reported, but there are inconsistencies, and the observed phenotype is not always supported by genotypic analysis. In particular, RB1 strains have been associated with virulence traits such as *in vitro* adherence to epithelial cells and serum resistance more frequently than RB2/3 strains, but genes that mediate these phenotypes have been found in strains from both lineages [64]. In addition to presence/absence studies on a specific gene or protein, it may also be necessary to examine sequence conservation
and allelic distribution. Some genes are found ubiquitously in the *M. catarrhalis* isolates examined, but allelic variants are differentially distributed. These differences are worth examining, as functional variation between alleles may contribute to differences between colonizing and disease-causing strains. In addition, it may also be necessary to investigate expression differences, as several virulence factors discussed in this review have phase-variable expression, and this may influence the course of infection as well as potential vaccine efficacy. Ideally, the assessment of whether factors are expressed should be determined from patient samples with minimal passaging (i.e. without subculture) if possible, or from a representative population of infectious bacteria (i.e. multiple colonies from subculture, to encompass the full range of expression differences present in the population). To date, few studies have considered these aspects when inferring relationships between isolates and clinical presentation.

Further complicating our understanding of virulence factor distribution is the lack of a standardized typing system that permits the comparison of worldwide isolates and associated epidemiological data (Box 1). The widely used typing systems for *M. catarrhalis* focus on whole-chromosome polymorphisms to distinguish different strains, however they provide limited information regarding the virulence of typed isolates due to their inability to assess the presence and allelic variation of individual virulence factors, and are difficult to compare between laboratories. An established *M. catarrhalis* MLST scheme is available that addresses the issue of ease of comparison between worldwide laboratories and provides an accessible reporting system, but this system is currently underutilized. Widespread implementation of this system would be invaluable for elucidating the population structure of *M. catarrhalis*, and could be coupled with a multi-virulence-locus sequence typing scheme to fully elucidate the virulence factor distribution of *M. catarrhalis*. An alternative to this is whole-genome sequencing of strains, but this is limited by cost constraints. Regardless of how it is achieved, a clearer understanding of how virulence determinants are distributed in the *M. catarrhalis* population would help to inform the selection of vaccine candidates.

The development of a *M. catarrhalis* vaccine could have far-reaching implications, reducing both infection by *M. catarrhalis* and co-infections synergized by *M. catarrhalis*. However, the selection of vaccine antigens is an ongoing process and numerous candidates are under evaluation. Strategic selection of vaccine antigens requires a full and clear understanding of the distribution of all candidates, and the assessment of their sequence variation and expression. The adoption of a universal typing system will provide a better understanding of the *M. catarrhalis* population structure and allow more accurate predictions to be made on the potential virulence determinant repertoire of uncharacterized isolates. It is hoped that the information obtained from such studies examining *M. catarrhalis* pathobiology will aid the development of an efficacious vaccine against *M. catarrhalis*.

Funding information
The authors are supported by the Australian National Health and Medical Research Council (NHMRC) (career development fellowship 1045235 and project grant 1099278 to K. L. S.), the Australian Government (APA to L. V. B), the Garnett Passe and the Rodney Williams Memorial Foundation (Fellowship to A. T.).

Conflicts of interest
The authors declare that there are no conflicts of interest.

References
7. Perez AC, Murphy TF. Do we need a vaccine against *Moraxella catarrhalis* in chronic lung disease? What are the options and opportunities? *Vaccine* 2017: Epub ahead of print.


