

# SCIENTIFIC REPORTS

OPEN

## Virulence factors of *Moraxella catarrhalis* outer membrane vesicles are major targets for cross-reactive antibodies and have adapted during evolution

Daria Augustyniak<sup>1</sup>, Rafał Seredyński<sup>2,3</sup>, Siobhán McClean<sup>4</sup>, Justyna Roszkowiak<sup>1</sup>, Bartosz Roszniowski<sup>1</sup>, Darren L. Smith<sup>5</sup>, Zuzanna Drulis-Kawa<sup>1</sup> & Paweł Mackiewicz<sup>6</sup>

*Moraxella catarrhalis* is a common human respiratory tract pathogen. Its virulence factors associated with whole bacteria or outer membrane vesicles (OMVs) aid infection, colonization and may induce specific antibodies. To investigate pathogen-host interactions, we applied integrated bioinformatic and immunoproteomic (2D-electrophoresis, immunoblotting, LC-MS/MS) approaches. We showed that OMV proteins engaged exclusively in complement evasion and colonization strategies, but not those involved in iron transport and metabolism, are major targets for cross-reacting antibodies produced against phylogenetically divergent *M. catarrhalis* strains. The analysis of 31 complete genomes of *M. catarrhalis* and other *Moraxella* revealed that OMV protein-coding genes belong to 64 orthologous groups, five of which are restricted to *M. catarrhalis*. This species showed a two-fold increase in the number of OMV protein-coding genes relative to its ancestors and animal-pathogenic *Moraxella*. The appearance of specific OMV factors and the increase in OMV-associated virulence proteins during *M. catarrhalis* evolution is an interesting example of pathogen adaptation to optimize colonization. This precisely targeted cross-reactive immunity against *M. catarrhalis* may be an important strategy of host defences to counteract this phenomenon. We demonstrate that cross-reactivity is closely associated with the anti-virulent antibody repertoire which we have linked with adaptation of this pathogen to the host.

*Moraxella catarrhalis* is an important human-restricted pathogen responsible for sinusitis and otitis media in children as well as infections of the lower respiratory tract, causing exacerbation of chronic obstructive pulmonary disease in adults<sup>1,2</sup>. The most promising vaccine candidates against *M. catarrhalis* have been shown to induce protective immunity in pulmonary clearance of the bacteria in animal models<sup>3,4</sup>. In human studies, the high antibody levels against few vaccine candidates were correlated with reduced bacterial carriage<sup>5</sup>. Nevertheless, there is currently no licensed vaccine against *M. catarrhalis*<sup>6</sup>. This species is thought to be composed of two distinct lineages, namely the more pathogenic sero-resistant (SR) strains and the less pathogenic sero-sensitive (SS) ones<sup>7,8</sup>. These two groups of strains show separate evolutionary histories and contain a distinct subset of lineage-specific genes, 33 and 49, respectively<sup>9</sup>. Some of the genes unique to the SR lineage are involved in phosphate metabolism but others do not have assigned functions. However, most of the characterized virulence factors are present in the

<sup>1</sup>Department of Pathogen Biology and Immunology, Institute of Genetics and Microbiology, University of Wrocław, Przybyszewskiego 63/77, 51-148, Wrocław, Poland. <sup>2</sup>Department of Physiology, Wrocław Medical University, T. Chalubinskiego 10, 50-368, Wrocław, Poland. <sup>3</sup>Department of Physical Chemistry of Microorganisms, Institute of Genetics and Microbiology, University of Wrocław, Przybyszewskiego 63/77, 51-148, Wrocław, Poland. <sup>4</sup>School of Biomolecular and Biomedical Sciences, UCD O'Brien Centre for Science West, B304, Dublin, Ireland. <sup>5</sup>Applied Sciences, University of Northumbria, Ellison Building EBD222, Newcastle upon Tyne, NE1 8ST, UK. <sup>6</sup>Department of Genomics, Faculty of Biotechnology, University of Wrocław, Joliot-Curie 14a, 50-383, Wrocław, Poland. Correspondence and requests for materials should be addressed to D.A. (email: [daria.augustyniak@uwr.edu.pl](mailto:daria.augustyniak@uwr.edu.pl)) or P.M. (email: [pamac@smorfland.uni.wroc.pl](mailto:pamac@smorfland.uni.wroc.pl))

core genome of both SR and SS strains. The difference in pathogenicity correlates with the clinical observation that most strains isolated from diseased patients possess the sero-resistant features<sup>10</sup>. The sero-resistant strains isolated from different patients are very similar with respect to their virulence, metabolic potential, CRISPR-cas, mobile genetic element content, gene content and chromosomal synteny<sup>11</sup>. The fact that nearly three-quarters of the genes are common, including those associated with virulence, indicates that a range of isolates of *M. catarrhalis* may be controlled by vaccination. A potential distinction between the isolates could result from an extremely small number of genes or epigenetic phenomena<sup>11</sup>. Such epigenetic regulation of multiple gene expression via the phase-variable DNA methyltransferase (ModM) regulon was recently reported for *M. catarrhalis* phase variations associated with otitis media<sup>12</sup>.

The mechanisms of colonization and pathogenesis of *M. catarrhalis* have been extensively studied and many virulence factors have been identified to date. The most important virulence strategies involve: (1) evasion of complement-mediated killing mainly via interference with regulatory proteins<sup>13–15</sup>; (2) polyclonal, non-specific B cell activation and redirecting of adaptive immunity<sup>16</sup>; (3) hiding inside lymphoid tissue, which is the main reservoir facilitating the host invasion<sup>17</sup>; (4) formation of biofilm<sup>18,19</sup>; and (5) participation in protease-antiprotease imbalance<sup>20</sup>. Some of these strategies can be driven in part by the release of outer membrane vesicles (OMVs), which contain several virulence factors facilitating the delivery of periplasmic and outer membrane components to the host<sup>14,21</sup>. Moreover, OMVs can favour pathogen coexistence and colonization after their interaction with the other bacterial species<sup>22,23</sup>.

Many immune-relevant and shared microbial epitopes stimulate the production of intraspecies and interspecies cross-reacting antibodies with high frequency and have cross-protective efficiency<sup>24,25</sup>.

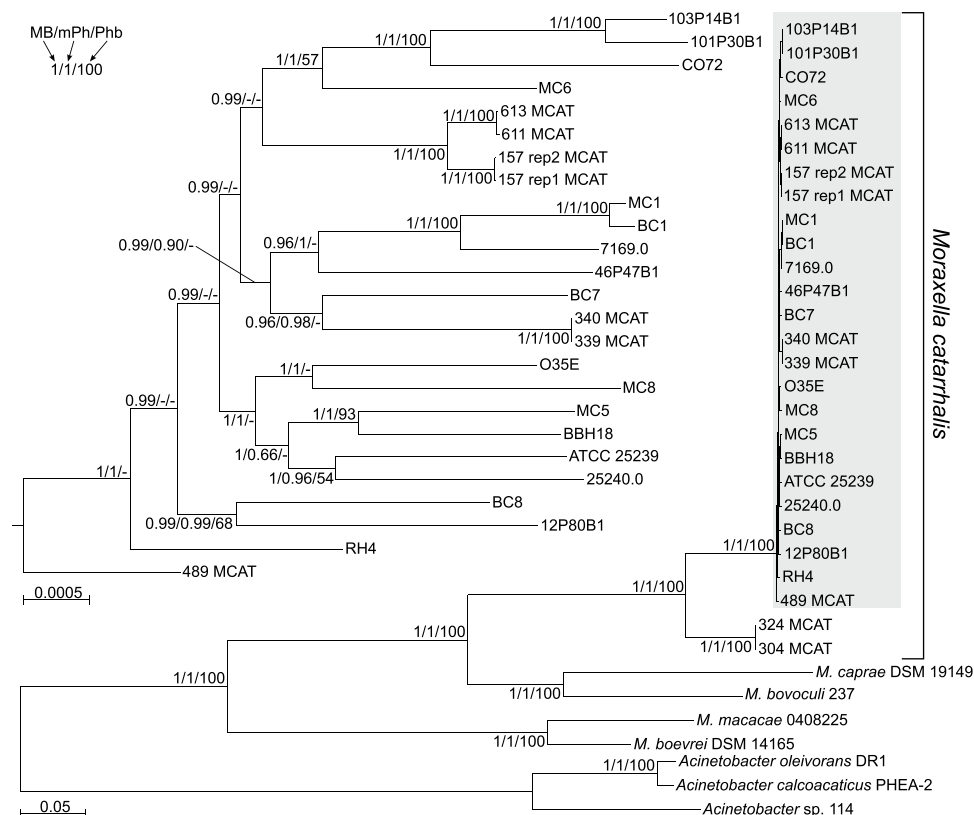
The major arm of defence against *M. catarrhalis* relies on antibody-dependent mechanisms<sup>6</sup>. We have previously reported that some of them, including bactericidal, opsonophagocytic and adhesion blocking protective function, can be driven *in vitro* by cross-reactive antibodies that primarily recognize outer membrane proteins (OMPs)<sup>26</sup>. To confer an anti-virulent strategy, the antibodies directed against *M. catarrhalis* may also act in concert with unique humoral mediators such as neuropeptides<sup>27</sup>. To investigate the nature and basis of our previous observations and define the specific profile of OMPs targeted by cross-reactive antibodies, we performed an immunoproteomic analysis of OMVs from two clinical strains of *M. catarrhalis* (Mc6 and Mc8) which differ in phenotypic properties, lipooligosaccharide (LOS) type, source of origin<sup>26</sup> and belong to distant phylogenetic lineages. Although a comprehensive proteomic analysis of OMVs from one *M. catarrhalis* RH4 strain was recently published<sup>28</sup>, immunoproteomic studies have not been performed to date. The immunoproteomic analysis of OMVs was characterized by 2D-electrophoresis associated with LC/MS mass spectrometry and carried out by cross-immunoreactivity experiments. Moreover, we performed extensive bioinformatics analyses of OMV proteins including four newly sequenced and all previously published *Moraxella* complete genomes to study features of these proteins from a genomic and evolutionary perspective with particular emphasis on their role in virulence.

## Results

**Phylogenetic relationship between the studied genomes.** To determine the evolutionary relationships of the newly sequenced genomes (Mc1, Mc5, Mc6, Mc8) in comparison to other *M. catarrhalis* strains (Supplementary Table S1), we performed phylogenetic analyses of 34 complete genomes using alignments of 739 orthologous groups of proteins (Fig. 1). Trees obtained by different methods show very similar topologies to trees based on the presence and absence of the orthologous groups (Supplementary Fig. S1). All *M. catarrhalis* strains create a fully supported monophyletic clade and exclude the four other *Moraxella* species isolated from animals. Apart from two *M. catarrhalis* strains, 324 MCAT and 304 MCAT, the other strains are very tightly grouped and are characterized by minimal divergence in comparison to other studied bacteria. The newly obtained strains, are distributed across various branches of the *M. catarrhalis* phylogenetic tree. Two *M. catarrhalis* strains, Mc6 and Mc8, selected for proteomic analysis, are placed in different clades separated by many other lineages, therefore representing relatively distant phylogenetic lineages considering intraspecies variations. Furthermore, both strains are phenotypically different in LOS type, autoagglutination ability, adherence potency, biofilm formation and source of origin as described previously<sup>26</sup>.

**Profiles of immunoreactive OMV proteins from *M. catarrhalis* strains.** The OMV protein profiles from the strains Mc6 and Mc8 were compared by 2-D electrophoresis (2D-SDS PAGE). The resulting peptide mass fingerprints were analysed using a *M. catarrhalis* protein data set including 18 proteins. Only minor variations in the patterns of resolved protein spots between the strains Mc6 and Mc8 were obtained with respect to relative concentration, charge and molecular weight of a few proteins (Table 1). Our proteomic studies identified ten additional OMV proteins to those already described by Schaar *et al.* (2011), which display molecular masses between 15 and 120 kDa (Supplementary Data S1). They represent different functional groups. Four of them are probably involved in binding and transport of iron ions: FbpA, MhuA, LbpB and TbpA. Others are glutamate dehydrogenase (GdhA), ubiquitous surface protein UspA1, structural protein (MipA), polyisoprenoid-binding periplasmic protein (YceI), molecular chaperone (GroEL) and dihydrolipoamide dehydrogenase (Lpd). With the exception of TbpA, identified in Mc6 OMVs (LC-MS/MS: MW 119.4, accession number SQ01\_05725, sequence coverage 34%, score 1734), nine of them were immunogenic (Table 1).

A pseudogene SQ02\_07885 showing similarity to the gene encoding lactoferrin binding protein (LbpB) was annotated in the Mc8 genome. Therefore, it has been later rechecked and verified by resequencing of 1.2 kb PCR product. The sequence of this truncated gene contains many stop codons and indels which disrupt its potential amino acid sequence and cause frame shifts. The first stop codon appears after 2214 nucleotides giving the potential amino acid product with the length of up to 738 residues. The original amino acid sequence would result in 906 residues. However, the truncated sequence contains a complete protein domain (pfam01298, Lipoprotein\_5, Transferrin binding protein-like solute binding protein) present in its full homologues. We confirmed that the



**Figure 1.** Phylogenetic relationships between *Moraxella* and *Acinetobacter* genomes. MrBayes tree based on the concatenated alignments of sequences from 739 orthologous groups represented 31 *Moraxella* genomes and three members of *Acinetobacter* genus. The part of tree marked by the grey rectangle is shown on the left in the larger scale. Numbers at nodes, in the order shown, correspond to: posterior probabilities estimated in MrBayes (MB), support values obtained by the approximate likelihood ratio test based on a Shimodaira-Hasegawa-like procedure (Anisimova and Gascuel O. 2006) calculated in morePhyML (mPh) and bootstrap values obtained in PhyML (Phb). Values of the posterior probabilities, SH-like branch supports and bootstrap percentages lower than 0.50 and 50%, respectively, were omitted or indicated by a dash “—”.

shortened protein retains its immunogenic properties. This protein product was identified in immunoblots and showed reactions with antisera as shown below.

**Immunoblot analysis of OMV proteins reveals a specific pattern of cross-reactivity with antibodies.** To study immunoreactive proteins present in OMVs from Mc6 and Mc8 strains, one homologous anti-serum and two cross-reactive, serotype-different murine sera were selected. These sera, namely anti-Mc1, anti-Mc6 and anti-Mc8 which were raised in response to whole bacteria in order to better mimic the natural reaction of the host, showed significant cross-protective bactericidal and anti-adhesive activity *in vitro*<sup>26</sup>. The characteristics of used antisera is shown in Table 2. The antisera were specific against strains representing separate lineages in the phylogenetic tree of *M. catarrhalis* strains (Fig. 1) and consequently represent a wide phenotypic variation within this species. The advantage of studying *Moraxella* cross-reactivity using mouse sera is connected with the fact that all antibodies specific against *M. catarrhalis* antigens in mouse are produced *de novo* without any interference from pre-existing antibodies.

In immunoproteomic studies, a combination of 2D-gel electrophoresis, western blot analysis and LC-MS/MS was performed. The spots that were immunoreactive with the cross-reactive and homologous antibodies were matched to the stained reference gel by comparison of their electrophoretic position, size and shape. The most immunoreactive proteins spanned a broad range of pI values between 4.5 and 10 and molecular mass from 20 to 80 kDa. Overall, 24 Mc6 and 27 Mc8 immunoreactive spots were successfully identified by mass spectrometry (out of 51 Mc6 and 52 Mc8 immunoreactive spots, respectively), representing 13 and 14 different proteins (Table 1). The number of immunoreactive spots in Mc6 and Mc8 was similar when probed with homologous antisera, up to 25% lower for more-potent heterologous antisera (anti-Mc6 or anti-Mc8) and up to 70% lower for less-potent heterologous antisera (anti-Mc1).

The analyses allowed the identification of a set of common proteins with different functions, mostly classified as outer membrane proteins, which were recognized by all three antisera used. Several of these immunoreactive proteins, including OMP CD were identified from more than one spot, often with the same molecular weight but with different pI values. These are likely to represent isoforms of the same protein that retain their immunogenicity. The analyses allowed also for an unambiguous identification of several proteins which showed

Protein description and name	Functional category <sup>1</sup>	MW (kDa) <sup>2</sup>	Strain	Accession number	Sequence coverage (%)	Score <sup>3</sup>
Outer membrane protein CD (OMP CD)	Cell envelope biogenesis, outer membrane <sup>1</sup> , cell binding, serum resistance	48.3	Mc6	SQ01_01315	23–56	328–1213
			Mc8	SQ02_01320	25–44	464–1064
Outer membrane protein E (OMP E)	Lipid metabolism <sup>1</sup> , fatty acid transport, serum resistance	49.3	Mc6	SQ01_04940	69–71	1473–1720
			Mc8	SQ02_04980	45–67	1105–1995
Outer membrane protein (CopB), TonB-dependent siderophore receptor	Inorganic ion transport and metabolism <sup>1</sup> , serum resistance	82.8	Mc6	SQ01_06490	30–58	1115–2204
			Mc8	SQ02_06550	33–50	1260–2019
Ubiquitous surface protein A (UspA1)	Cell division and chromosome partitioning <sup>1</sup> , cell binding, serum resistance	93.3	Mc6	SQ01_03435	24–41	596–1011
			Mc8	SQ02_03470	22–23	787–782
Outer membrane porin (OMP M35)	Cell envelope biogenesis, outer membrane <sup>1</sup> , transport and binding proteins	38.0	Mc6	SQ01_03260	38	533
			Mc8	SQ02_03275	78	1286
Transferrin binding protein B (TbpB)	transport and binding proteins	76.2	Mc6	SQ01_05710	22–38	884–1140
			Mc8	SQ02_05755	21	710
Lactoferrin binding protein (LbpB)	Transport and binding proteins	100.1	Mc6	SQ01_07800	10	353
			Mc8	SQ02_07885	13	558
Haemoglobin-utilization protein, TonB-dependent receptor, (MhuA)	Inorganic ion transport and metabolism <sup>1</sup> , transport and binding proteins	107.6	Mc6	SQ01_05485	37–40	1422–1460
			Mc8	SQ02_05535	ND	ND
Structural protein (MipA)	Cell envelope biogenesis, outer membrane <sup>1</sup>	28.6	Mc6	SQ01_04155	53–60	498–577
			Mc8	SQ02_04180	75	799
Fe(3+) ABC transporter substrate-binding protein (FbpA)	Inorganic ion transport and metabolism <sup>1</sup>	35.7	Mc6	SQ01_07425	ND	ND
			Mc8	SQ02_07515	58–73	849–1080
Molecular chaperone (GroEL)	Posttranslational modification, protein turnover, chaperones <sup>1</sup>	57.4	Mc6	SQ01_02125	ND	ND
			Mc8	SQ02_02135	60	1721
Dihydrolipoamide dehydrogenase (Lpd)	Energy production and conversion <sup>1</sup>	51.1	Mc6	SQ01_00050	29	737
			Mc8	SQ02_00050	46	1204
Hypothetical protein with TPR domain (HP Mc4)	General function prediction only <sup>1</sup>	30.1	Mc6	SQ01_03330	46	554
			Mc8	SQ02_03345	62	758
Outer membrane beta-barrel protein (OMPJ)	Cell envelope biogenesis, outer membrane <sup>1</sup>	18.9	Mc6	SQ01_07075	ND	ND
			Mc8	SQ02_07135	34	246
Polyisoprenoid-binding periplasmic protein (YceI)	General function prediction only <sup>1</sup>	21.2	Mc6	SQ01_05855	ND	ND
			Mc8	SQ02_05915	63	528
Membrane protein (OMP G1b)	Cell envelope	26.9	Mc6	SQ01_02650	29	448
			Mc8	SQ02_02660	ND	ND
Membrane protein, Opa-like protein A (OlpA)	Cell envelope biogenesis, outer membrane <sup>1</sup>	23.3	Mc6	SQ01_01010	50	296
			Mc8	SQ02_01010	ND	ND

**Table 1.** Immunogenic proteins determined by LC-MS/MS of excised spots found by immunoblotting on 2DE gels. <sup>1</sup>Functional category according to search results of Cluster of Orthologous Groups (COGs); other descriptions were taken from various references; <sup>2</sup>Theoretical molecular mass was determined by Mascot; <sup>3</sup>Score ranges represent MS/MS ion scores determined by peptide mass fingerprinting. Only scores deemed to be significant by Mascot ( $p < 0.05$ ) were presented. ND - not done as there was no reactivity observed on blots.

qualitative and quantitative differences in immunoreactivity. As presented in Figs 2 and 3, the most immunoreactive -cross-reacting proteins showing the highest staining intensities for IgG antibodies seem to be the outer membrane proteins involved in bacterial adhesion (OMP CD, UspA1) or complement resistance (UspA1, OMP E, OMP CD, OMP J/OlpA/OMP G1b). On the other hand, the predicted proteins involved in iron acquisition such as TbpB or LbpB, including the truncated version of LbpB in Mc8, were rather weakly or not recognized by cross-reactive antibodies, while their recognition by homologous sera was abundant for the relevant studied strain. A similar pattern of reactivity was observed for protein engaged in iron acquisition by hemoglobin utilization such as MhuA. Iron-binding protein CopB gave a variant and strain-dependent pattern of cross-reactivity. The lack of cross-reactivity was noticed for periplasmic ferric binding protein FbpA abundantly expressed in Mc8. There were also a few proteins predicted to be both cytoplasmic and surface exposed. Among them were Lpd and GroEL chaperone which showed moderate immunoreactivity in both strains while probing with anti-Mc6 or anti-Mc8 sera in heterologous manner. Taken together, the more-potent cross-reactive antisera (anti-Mc6 or anti-Mc8) recognized 11 different proteins with high or moderate spot intensities, whereas the less potent antiserum (anti-Mc1) recognized between 5 and 7 different proteins (depending on the strain used) thus showed lower intensities of reactions in general. The original full-length images of 2D-gels and blots can be seen in Supplementary Fig. S2.

Serum	Strain Mc6			Strain Mc8		
	Titer <sup>1</sup>	Avidity index <sup>2</sup>	Bactericidal titer <sup>3</sup>	Titer <sup>1</sup>	Avidity index <sup>1</sup>	Bactericidal titer <sup>3</sup>
anti-Mc1	246,254	46 ± 4.2	261	255,242	34 ± 0.7	75
anti-Mc6	2,140,484	79 ± 1.4	607	751,395	53 ± 1.4	317
anti-Mc8	255,290	57 ± 0.7	234	1,001,067	43 ± 2.8	720
pre-immune	<500	ND	<3	<500	ND	<3

**Table 2.** Characteristics of murine antisera. <sup>1</sup>The end-point of IgG ELISA titer against OMVs is defined as reciprocal value of the antiserum dilution giving an absorbance of control non-immunized mouse serum enlarged by two standard deviations (2 SD). The values are presented as geometric mean titers from two independent experiments. <sup>2</sup>Avidity was tested by OMVs ELISA elution assay, using a chaotropic agent 1.5 M sodium thiocyanate (NaSCN) to disrupt weak antigen-antibody binding. The results are expressed as mean avidity index (AI) with SEM where  $AI = (OD_{490} \text{ with NaSCN} / OD_{490} \text{ without NaSCN}) \times 100$ . <sup>3</sup>The titer was defined as the reciprocal value of the highest dilution of antiserum at which  $\geq 50\%$  killing of the target strain was observed. The results were published in our previous paper<sup>26</sup>.

Due to the size limitations of 2D electrophoresis for the certain high molecular weight (HMW) proteins of *M. catarrhalis* such as Hag/MID or UspA2, their presence in OMVs and their reactivity with antisera were performed by classical SDS-PAGE and Western blot. Strong cross-reactivity was only observed for the ~250 kDa HMW antigen (presumably Hag/MID) of Mc6 and Mc8. There was no cross-reactivity for the other HMW protein, which was abundantly visible on the gel, with a MW in excess of 300 kDa. OMVs from Mc6 but not Mc8 *Moraxella* strains heated at 37 °C and 100 °C for 10 min showed the presence of this HMW antigen, which indicates the predicted presence of UspA2 antigen with very high molecular size and a multimeric structure that was not reduced by heating at 100 °C<sup>29</sup>. The results for HMW antigens are presented in Supplementary Fig. S3.

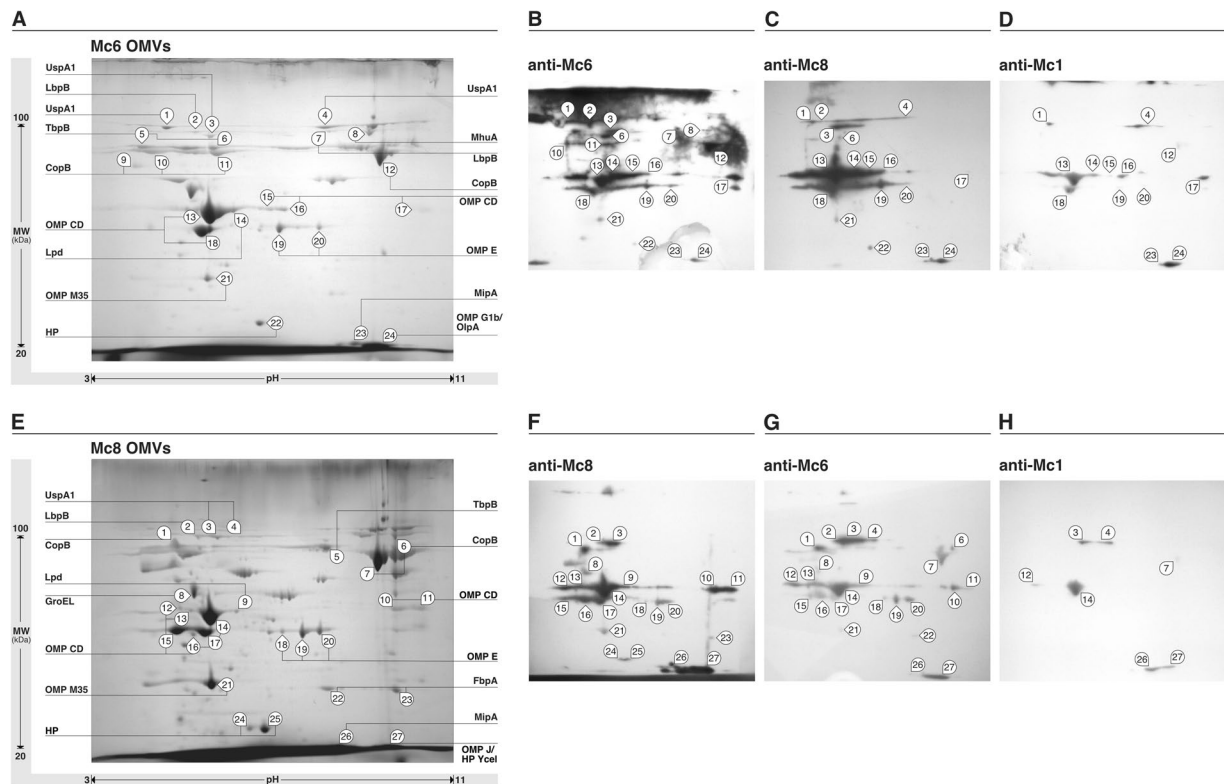
**Potential epitopes in OMV proteins subjected to cross-reaction studies.** It is interesting to hypothesize on the reasons why some proteins were able to react with antisera elicited by different *Moraxella* serotypes while others were not. One could assume that it is related to various level of sequence conservatism of these proteins. Proteins with lower sequence divergence should respond to the cross-reactions with greater probability and intensity than more divergent ones. Indeed, the homologous proteins involved in strong positive cross-reactions were characterized on average by a significantly smaller mean percentage difference between their entire sequences (2.85%) than the proteins engaged in weak or no cross-reactions (8.04%), ( $p$ -value = 0.04). However, the global sequence divergence may not well reflect the variation/conservatism of epitopes responsible for the cross-reactivity.

Therefore, we searched the sequences of proteins for potential epitopes using 15 bioinformatic prediction methods. As potential epitopes, we selected sites that were recognized by more than half of the approaches. Plots showing the prediction of epitope sites along sequences of the studied proteins are presented in the Supplementary Fig. S4. Supplementary Data S2 includes the list of individual amino acid residues that were predicted to be involved in the epitopes of individual proteins. The fractions of the sites predicted as epitopes were quite variable (Supplementary Data S3). Assuming the 50% prediction threshold, the percent of these sites in sequences range from 0.4% (Lpd protein from Mc1 and Mc6 strains) to 22.1% (LbpB protein from Mc8 strain). Proteins that were not involved in strong cross-reactions had a significantly larger contribution of predicted epitope sites in their sequences (on average  $9.9 \pm 6.2\%$  SD), than the proteins engaged in such cross-reactions ( $6.1 \pm 3.8\%$  SD,  $t$ -test,  $p$ -value = 0.012). The predicted epitope sites of homologous proteins engaged poorly or not at all in strong cross-reactions were also characterized on average as having almost twice the percentage differences between their sequences (12.9%) as the epitope sites of homologous proteins involved in the cross-reactions (6.8%).

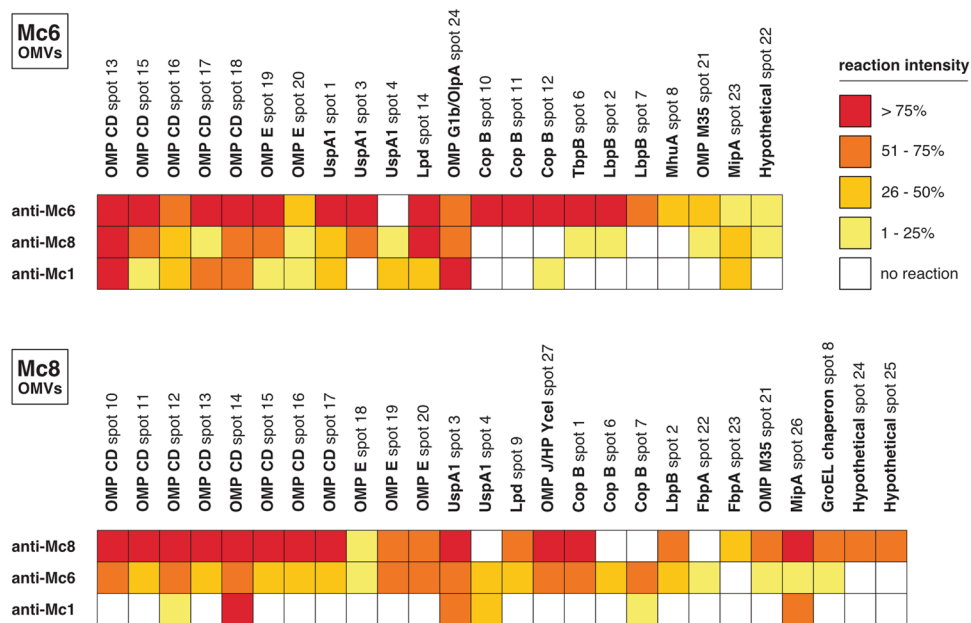
**Distribution of OMV proteins in the studied genomes.** To study and characterize OMV proteins in a more general functional and evolutionary context, we first clustered all protein sequences from four newly sequenced and 23 publicly available full genomes of *M. catarrhalis* as well as four other *Moraxella* species into orthologous groups (Supplementary Table S2). Three representatives of *Acinetobacter* (family *Moraxellaceae*) were added as reference genomes because it is the most closely related genus (with the published complete genome) to *Moraxella*. The analyses revealed 4780 orthologous groups represented by at least two of the 34 studied genomes and 779 groups had at least one protein in each of the genomes (Supplementary Data S4).

OMV proteins were assumed to be those that were previously identified in proteomic studies by Schaar *et al.* (2011) and by us. Using homology searches, we assigned the OMV proteins to the orthologous groups. These proteins were recognized in 64 orthologous groups (Supplementary Data S1), which contained 1995 annotated proteins in total from the studied genomes. Among the 64 OMV protein groups, 43 were conserved and present in at least one representative of *Acinetobacter* as well as in *M. catarrhalis* and other *Moraxella* species (Fig. 4A), while 24 were found in all genomes studied. However, a substantial proportion (25%), i.e. 16 of the OMV protein groups were represented only in *Moraxella* and five OMV protein groups were restricted to *M. catarrhalis*: IclR (W2387), Lgt1 (W2401), OppB (W2484), cHP Mc10 (W2552) and HP Mc12 (W2666) (Supplementary Data S1). Interestingly, no such unique OMV proteins were found in *Acinetobacter* or in other *Moraxella* species although their genomes encode many unique proteins of other types (Fig. 4B). A likelihood ratio test (G-test) of independence showed significant differences between the distribution of orthologous group for total and OMV proteins ( $p$ -value < 2.2E-16).

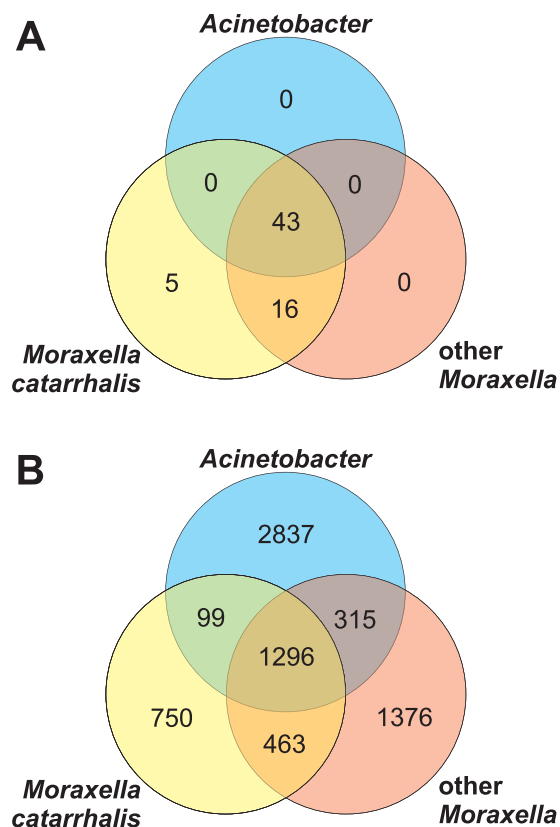




**Figure 2.** Immunoproteomic analyses of *M. catarrhalis* OMV preparations. After 2D-gel electrophoresis OMV proteins were visualized by Coomassie staining (A,E). 2D immunoblots probed with homologous antisera (B,F), stronger cross-reactive antisera (C,G) and weaker cross-reactive antisera (D,H). Molecular size markers are indicated on the left. The exposure times for chemiluminescent substrate were exactly the same for each presented blot. Gel and blot images are cropped and brightness/contrast-adjusted for better readability. Original images are shown in Supplementary Fig. S2.



**Figure 3.** Heatmap of immunoreactive spot intensities. Reactivity of murine post-immunization sera with OMVs proteins was estimated semi-quantitatively on a quartile scale. Only proteins that reacted with at least one antisera are shown.



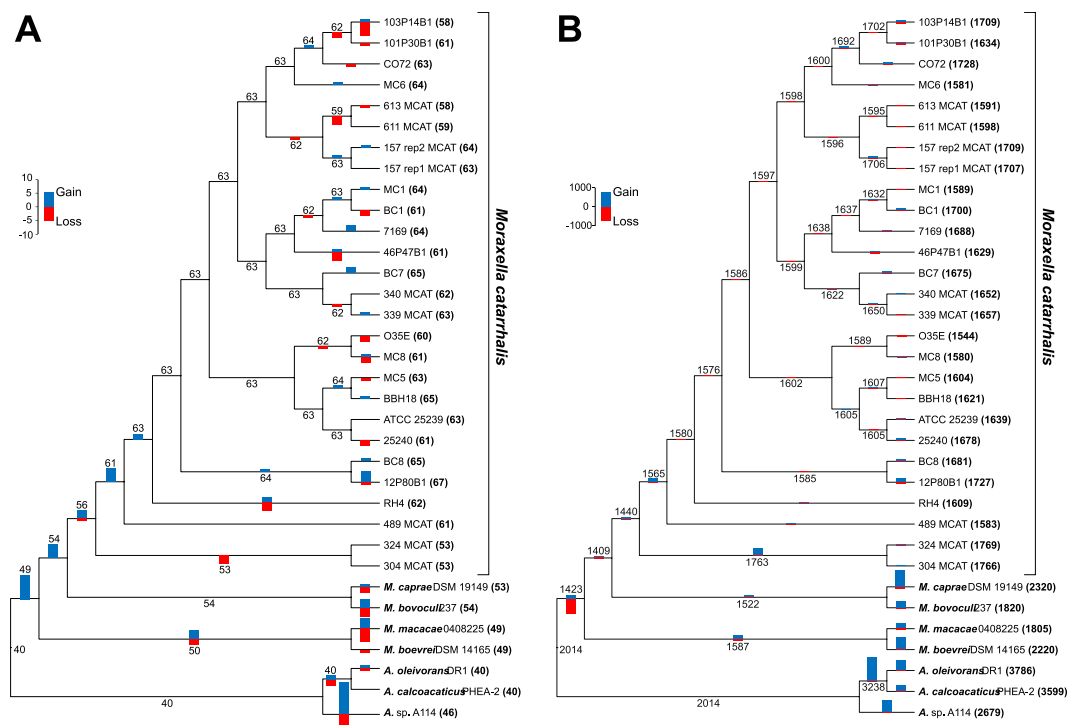
**Figure 4.** Distribution of orthologous group for OMV proteins (**A**) and total proteins (**B**) found with OrthoMCL across three sets of genomes: *Acinetobacter*, *Moraxella catarrhalis* and other *Moraxella* species. The group was included if it was represented by at least one genome in the set. The unique proteins not classified into groups were also included.

Forty four OMV protein groups were found in all 27 *M. catarrhalis* strains and 32 were identified in all 31 *Moraxella* strains examined. Homologues to hypothetical protein MCR\_0740 were the least represented in the genomes because they were found in only 10 *M. catarrhalis* strains. Genes coding for a phage-related protein PmpL and ubiquitous surface protein A UspA1 were subjected to duplication in several *M. catarrhalis* strains.

**Functional categories of OMV proteins.** The majority of all identified OMV proteins were grouped into three functional categories: (1) Cell wall/membrane/envelope biogenesis, (2) Translation, ribosomal structure and biogenesis and (3) Inorganic ion transport and metabolism. Each of these categories contained more than 10% OMV proteins. In turn, the functional groups most enriched in the OMV proteins in comparison to non-OMV proteins were: Secondary metabolites biosynthesis, transport and catabolism; Cell wall/membrane/envelope biogenesis; Cell cycle control, cell division, chromosome partitioning/Cytoskeleton and Extracellular structures. They were represented twice as much among OMV proteins than non-OMV proteins (Supplementary Table S3). It is also interesting to compare distribution of total proteins into functional categories across *M. catarrhalis* strains infecting human upper respiratory tract and *Moraxella* isolated from animals (Supplementary Table S4). For many categories, the differences between the fractions were statistically significant. The human pathogens had over twice as many proteins involved in: Amino acid transport and metabolism; Defence mechanisms; Nucleotide transport and metabolism and Coenzyme transport and metabolism.

**Changes in content of genes for OMV proteins in the studied genomes.** Using the orthologous group distribution and the phylogenetic tree of the studied genomes, we estimated sizes of gene content in ancestral genomes as well as gains and losses of genes in individual phylogenetic lineages. We compared the results of these studies between genes encoding OMV proteins and all other annotated protein sequences (Fig. 5).

The size of the ancestral OMV protein set for the common ancestor of *Moraxella* and *Acinetobacter* was assessed to be 40 (Fig. 5A). This number remained almost unchanged during the evolution of *Acinetobacter* genomes because they encode currently 40 or 46 such proteins. In contrast, during the evolution of *Moraxella*, the number of genes encoding these proteins continuously increased. Its successively diverging lineages comprised an increasing number of these genes: 50 (the common ancestor of *M. boevrei* and *macacae*), 54 (the common ancestor of *M. bovoculi* and *caprae*), 56 (the common ancestor of all *M. catarrhalis* strains), 61 (the common ancestor of *M. catarrhalis* strains excluding the basal 324 MCAT and 304 MCAT strains). The subsequent increase in this gene content resulted in 58 to 67 such genes per genome in *M. catarrhalis* strains that evolved after separation of



**Figure 5.** Gains and losses of genes for OMV proteins (A) and all annotated protein sequences (B) during evolution of *Moraxella* and *Acinetobacter* genomes. The number of gene gains and losses per phylogenetic lineage were presented by bars. Ancestral gene contents in genomes were shown by numbers at tree branches and the current gene contents were presented by numbers in parenthesis at individual genomes.

the 324 MCAT and 304 MCAT strains. Among the evolved genomes, only small fluctuations in the gene content has occurred: up to 5 losses (the lineage of 103P14B1) and up to 4 gains (the lineage of 1280B1 strain).

The increase in the number of genes coding for OMV proteins does not seem to be an artefact related to the homology searches in the closely related genomes of *M. catarrhalis* versus the others because the same analysis performed with all protein-coding genes did not reveal the trend typical of the genes for OMV proteins. Furthermore, the whole proteome showed the opposite trend (Fig. 5B). In contrast to the content of genes encoding OMV proteins, ancestors of *Moraxella* decreased their global protein-coding gene content by almost 600 genes, whereas *Acinetobacter* genomes increased their genomes by 665 to 1772 genes in comparison to their common ancestor with 2014 genes. Some increase in the global gene content in *Moraxella* arose much later during evolution and rather independently in individual lineages. The most gene-abundant genomes evolved in *Moraxella* isolated from animals (1805 to 2320 genes), while human pathogenic *M. catarrhalis* strains encode only 1544 to 1769 proteins.

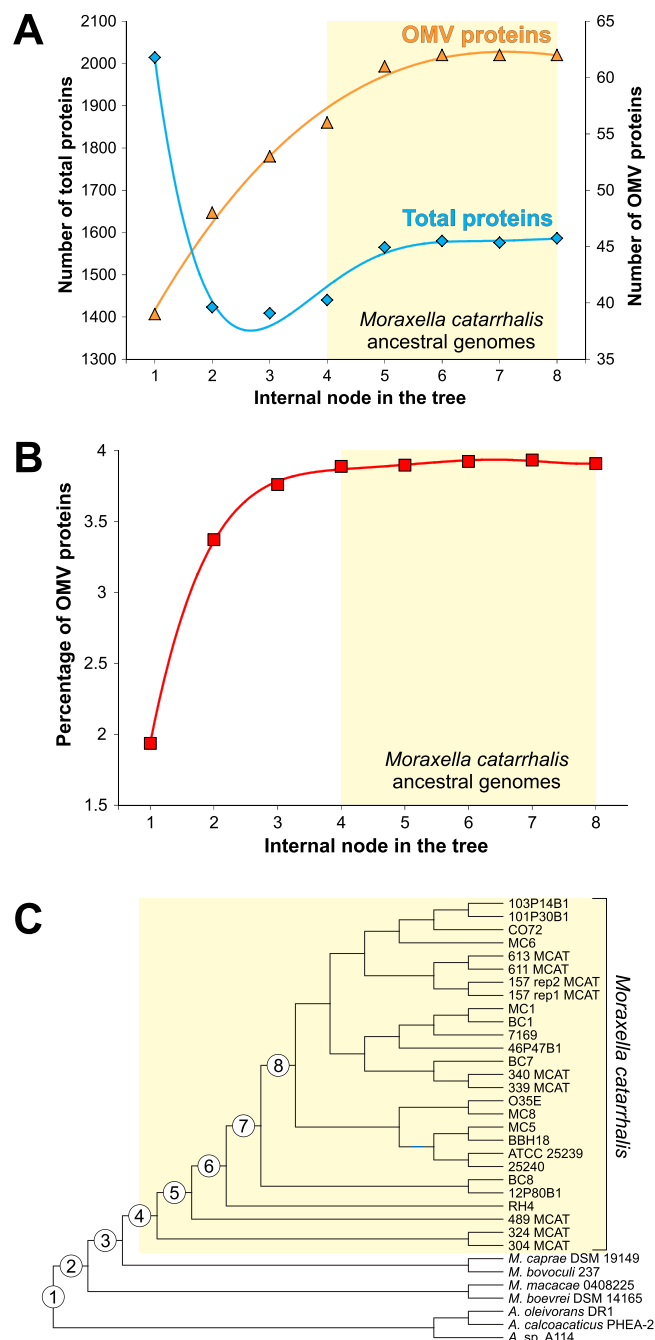
These trends in gene content alteration are apparent in Fig. 6A, where the numbers of genes for total or OMV proteins estimated for ancestral genomes were plotted along the subsequent internal nodes of the phylogenetic tree (Fig. 6C). The number of OMV protein genes clearly increased during evolution and reached a stable maximum for *M. catarrhalis* ancestral genomes. Simultaneously, the total number of protein-coding genes rapidly dropped and subsequently increased only slightly. As a result of these tendencies, the percentage of OMV proteins coded in the ancestral genomes doubled, from about 2% to almost 4% (Fig. 6B).

Since OMV proteins are common antigens, it was interesting to establish if their sequences showed a greater variation than other types of proteins. The median percentage difference of the OMV proteins was slightly larger than that of the other proteins: 16.2% and 15.4%, respectively.

## Discussion

Proteins associated with OMVs are engaged in diverse biological processes enabling infection and colonization of host by pathogens. Many of them are important virulence factors enhancing the potential to cause diseases<sup>30</sup>. Our comprehensive computational analyses of 31 complete *Moraxella* genomes and three representatives of *Acinetobacter* showed that OMV proteins identified in *M. catarrhalis* are organized in 64 orthologous groups and involved in various functions, especially those related to cell membrane and inorganic ion transport. Interestingly, we found that *M. catarrhalis* genomes are significantly enriched in these OMV proteins in comparison to other *Moraxella* species isolated from animals. It indicates that these proteins are not conserved across other taxa and have been quite recently acquired by *M. catarrhalis*. The acquisition of their genes can be associated with the emergence of these human-specific pathogens and is confirmed by the leading role of OMV in the increasing infectivity, pathogenicity and virulence of these bacteria. It has been documented that proteins incorporated in *M. catarrhalis* OMVs can play a role in complement evasion, epithelial adhesion, redirection of adaptive





**Figure 6.** Changes in gene content during evolution of *Moraxella* and *Acinetobacter* genomes. Number of genes for total proteins and OMV proteins estimated for ancestral genomes (A) in subsequent internal nodes from 1 to 8 in the phylogenetic tree presented in (C). Percentage of OMV proteins coded in the ancestral genomes from subsequent internal nodes (B).

humoral response, beta-lactam antibiotic cleavage and biofilm formation<sup>16,23,28,31</sup>. On the other hand, the lack of OMV enrichment in *Moraxella* species infecting other mammals could suggest either a major involvement of other virulence factors (e.g. cytotoxins or type IV pili in *M. bovis* and *M. bovoculi*) or alternatively their lower pathogenicity<sup>32,33</sup>.

Since OMV components in general are strong immunomodulatory antigens, they can significantly influence the mammalian immune system<sup>21,34</sup> and stimulate it to develop effective cross-protective defence<sup>35,36</sup> which in turn may influence the outcome of natural infections<sup>37,38</sup>. Recently, we have reported the wide range of effector anti-virulent functions fulfilled by cross-reacting antibodies against phenotypically different *M. catarrhalis* clinical strains<sup>26</sup>. In contrast to the previous proteomic studies on OMVs from RH4 strain of *M. catarrhalis*<sup>28</sup>, our present approach, focused on immunoproteomic analyses, revealed a particular set of antigens associated with virulence and cross-immunity that showed a similar pattern of responsiveness. The antigens that were highly

cross-recognized by sera including OMPCD, UspA1, OMPE, OlpA/OmpJ and presumably Hag/MID, were specific proteins responsible for serum resistance/complement evasion and/or adhesion enhancement. Therefore, the basis of antibody-mediated action directed against heterologous *M. catarrhalis* strains seems to be multidirectional blocking of these pivotal virulence factors. The observed significant response of cross-reacting antibodies to cytosolic/surface exposed Lpd and in part also to GroEL chaperone seems to be in line with this, as it reveals quite potent immunogenic properties of proteins whose virulent potential is well described. To our knowledge, this is the first study showing the immunogenicity of Lpd. In addition, the involvement of *Pseudomonas* Lpd in inhibition of complement cascade has been demonstrated<sup>39</sup>. GroEL immunoreactivity was observed in *Burkholderia*<sup>40</sup> infections. Likewise, the immunogenicity and protective efficacy of GroEL was confirmed for *Streptococcus pneumoniae* against a lethal infection in mice<sup>41</sup>. Although the surface exposition of GroEL and Lpd have not been directly confirmed to date for *M. catarrhalis*, their presence on the surface of other bacteria including *Neisseria meningitidis*<sup>42</sup> and *Pseudomonas aeruginosa*<sup>43</sup> suggests a similar location. We were not able to detect UspA2 protein involved in serum resistance on blots, despite the confirmed presence of a heat-resistant HMW protein (size exceeds 300 kDa on the gel) at least for Mc6. Its identity is implied due to it being highly resistant to boiling, a biochemical feature of UspA2. The lack of immunoreactivity for UspA2 may be a result of phase variation<sup>29</sup> or poor transfer to blotting membranes as in case of other bacterial HMW proteins<sup>44</sup>.

The apparent presence of specific cross-reactive patterns (the qualitative cross-reactive threshold) was next confirmed for a distinct group of abundantly expressed OMV antigens, involved in iron acquisition from Fe-sequestering host proteins. This group contained protein antigens which did not appear to be classified as pivotal virulence factors and were not recognized, or only weakly recognized, by cross-reacting antibodies. Interestingly, three of them, namely transferrin-binding protein (TbpB) and lactoferrin-binding protein (LbpB), as well as CopB (responsible for binding of both transferrin and lactoferrin) which gave strong homologous reaction in our study, have been reported to induce an antibody response in convalescent sera or sputum<sup>1,45</sup>. It supports their specific rather than cross-reactive recognition. We found no reactivity with TbpA and LbpA proteins encoded by two *Moraxella* strains. There is still no conclusive evidence for antibody against *M. catarrhalis* TbpA or LbpA in convalescent phase sera<sup>45</sup>, although such results exist for experimental meningococcal infection<sup>46</sup>. The inefficiency in TbpA refolding during the electroblotting procedure also cannot be ruled out. For others such as hemoglobin-binding protein (MhuA)<sup>47</sup> or FbpA, responsible for iron transport from periplasmic space into the cell interior<sup>48</sup>, there is no data on their immunogenic significance to our knowledge. Although, the induction of functional neisserial anti-FbpA antibodies in an animal model has been described<sup>49</sup>, their bactericidal importance has not been confirmed<sup>50</sup>.

We found that epitope sequences of homologous non-virulent OMV proteins showing either weak or no cross-reactivity were more variable than those of virulent proteins involved in the strong cross-reactions. It clarifies why antibodies prepared on one strain cross-reacted with the other strain, most probably with the conserved epitopes. Therefore the quantitative increase in OMV protein content during evolution observed in this study could compensate the lack of any qualitative changes expressed by conservatism at the sequence level of many protein epitopes. Lack, or presence, of cross-reactivity may be due to amino acid replacements or deletions/insertions in the sequences. For example, we found sites in the sequence of TbpB proteins that are well recognized as epitopes in Mc6 strain by 11 and 12 methods but in strains Mc1 and Mc8 by much smaller number of the methods (Supplementary Fig. S5). The difference in the prediction is associated with substantial changes in their sequences. Consistent with this, the proteins from the Mc6 strain showed practically no cross-reactivity with antisera prepared from the Mc1 or Mc8 strains. However, mutations cannot explain the complete variation in cross-reactivity, because a few proteins that have identical sequences were not recognized by the cross-reactive antibodies, e.g. FbpA from Mc1 and Mc8 strains, GroEL from Mc1 and Mc6 strains and OMP M35 from all three strains. This observation implies than other factors can influence the cross-reactivity, e.g. posttranslational modifications, different exposition of protein in OMVs, intrinsic features of antigenic epitope such as its accessibility and density or at least insufficient levels of certain cross-reactive antibodies<sup>51–53</sup>.

Taken together, the majority of cross-reactive antibodies was directed against several *M. catarrhalis* targets that are involved in serum resistance and/or promotion of bacterial attachment and as such contribute to the pathogenicity of bacteria. It is also difficult to judge their individual importance *in vivo*. The advantage for the host in having antibodies with a range of specificities against bacterial antigens associated with serum resistance is to facilitate an effective defence even if particular sets of (monovalent) antibodies are not present in the antiserum.

In conclusion, the cross-reactive humoral immune response elicited by the whole bacteria is directed against *M. catarrhalis* antigens engaged exclusively in complement evasion and colonization/invasion strategies but not those which participate in nutrient or compound acquisition. Despite the fact that strain-specific immune responses are a hallmark of many bacterial infections including those caused by *M. catarrhalis*<sup>34</sup>, the immunity based on cross-reacting anti-virulent activity of antibodies<sup>26</sup> is likely to be an additional source of defence, which increases the likelihood of successful eradication of this human pathogen. The significant pattern of cross-reactivity against pivotal virulence targets is therefore advantageous for the host. On the other hand, cross-reactive immunity may have also significant evolutionary implications for bacteria virulence strategies. The observed twofold rise in the number of specific OMV proteins during evolution of *M. catarrhalis* is an interesting example of pathogen adaptation in successful colonization of the host and increasing virulence of this species. Our findings suggest that virulence factors associated with OMVs and cross-reactive antibodies are pivotal components of *M. catarrhalis*-host interaction and may have implications in the host-pathogen interplay. Further *in vivo* studies are needed to confirm the role of *M. catarrhalis* OMVs in cross-protection.

**Limitations of the data.** The conclusions of our study are based on results performed with murine antisera developed to antigens of whole bacteria and with previously documented *in vitro* protective capacity with

human cells. The immunoproteomic component of our study is therefore based on common physicochemical features shared by all protein surfaces including those engaged in antigen-antibody interactions. Furthermore, the advantage of studying *Moraxella* cross-reactivity using mouse sera is connected with the fact that all antibodies specific against *M. catarrhalis* antigens in mouse are produced *de novo* without any interference from pre-existing antibodies. Nevertheless, since *M. catarrhalis* is a human specific pathogen, limitations of the mouse models have to be taken into account. In this context the relevance of our study to natural infections of human host should be confirmed in order to verify that our observed results mimic the human host response.

## Material and Methods

**Reagents.** Acetonitrile (Sigma), Antibodies: peroxidase (HRP)-conjugated affinity pure goat anti-mouse IgG Fc $\gamma$ -specific, (Jackson ImmunoResearch); bovine serum albumin BSA (SERVA); Immobilon Western Chemiluminescent HRP Substrate (EMD Millipore Inc.), sodium thiocyanate NaSCN (Sigma); sequencing grade modified porcine trypsin (Promega), TMB substrate reagent (R&D).

**Bacterial strains and culture.** Two *M. catarrhalis* strains were used in the experimental studies: Mc6 and Mc8. The bacterial strains were routinely cultivated in BHI (brain heart infusion) broth at 37 °C with orbital shaking (150 rpm) and their characteristics were described previously<sup>26</sup>.

**Outer membrane vesicles isolation.** Outer membrane vesicles (OMVs) were isolated according to Schaar *et al.*<sup>28</sup> with some modifications. Briefly, 18 h cultures of *M. catarrhalis* strains were diluted 50-fold in 500 ml Brain Heart Infusion (BHI) media and incubated at 37 °C for 16–18 h, shaking (150 rpm). The cultures were harvested by centrifugation (8000 rpm for 15 min at 4 °C). The supernatants were collected and passed through 0.22  $\mu$ m pore size filter vacuum pump (Merck, Millipore). The filtrates were concentrated using 50 kDa Vivaspin centrifugal concentrators (Amicon ultra, Millipore) at 5000  $\times$  g for 30 min at 4 °C. The concentrated supernatants were subsequently pelleted overnight (100 000  $\times$  g, at 4 °C) in an ultracentrifuge (Beckman Coulter Optima). The pellets containing OMVs were re-suspended in 500  $\mu$ l of sterile PBS buffer (pH 7.4), aliquoted and stored in –20 °C. The sterility of the OMV preparations was confirmed on BHI agar. The protein concentrations in OMVs preparations were measured using Qubit fluorimeter or Bradford assay (Sigma-Aldrich Inc.) and the quality of OMVs preparations was confirmed in 12% SDS-PAGE.

**2-D gel electrophoresis of OMV proteins.** Two-dimensional electrophoresis (2-DE) procedure was carried out as previously described<sup>40</sup>. Samples were solubilized in rehydration solution (8 M urea, 2 M thiourea, 4% (w/v) CHAPS, 1% (v/v) Triton X-100, 10 mM Tris, 65 mM DTT, 0.8% (v/v) IPG buffer (pH 3–11, GE Healthcare Inc.), 0.01% (w/v) bromophenol blue) to the final protein concentration of 1 mg/ml. IPG strips (7 cm, pH 3–11NL, GE Healthcare Inc.) were saturated overnight with 120  $\mu$ l of this solution and subjected to isoelectric focusing in a GE Healthcare Ettan IPGphor 3 Isoelectric Focusing Unit (3 hours, total voltage of 7000 V). Subsequently, IPG strips were equilibrated for 20 minutes in reducing buffer (6 M urea, 50 mM Tris, 30% (v/v) glycerol, 2% (w/v) SDS, 2% (w/v) DTT) followed by 20 minutes in alkylating buffer (6 M urea, 50 mM Tris, 30% (v/v) glycerol, 2% (w/v) SDS, 2.5% (w/v) iodoacetamide). The second dimension was resolved on 12% 1 mm-thick SDS-PAGE mini gels, with the separation voltage of 130 V. For each experiment, a set of three or four identical gels was prepared. One gel was stained with Page Blue Coomassie staining solution (Fermentas Inc.) according to the manufacturer's instructions. The other gels were prepared for the further immunoblotting.

**Immunoblotting.** Gels were saturated in transfer buffer (25 mM Tris, 192 mM glycine, 20% (v/v) methanol) for 15 minutes. Proteins were transferred to PVDF membranes with a semi-dry Transphor unit (BioRad Inc.) at 330 mA for 50 minutes. PVDF membranes were subsequently blocked with 5% BSA solution in the TBST buffer (20 mM Tris-HCl, 150 mM NaCl, 0.05% (v/v) Tween 20, pH 7.5) for 1 hour at room temperature. The blots were probed overnight at 4 °C with post-immunized pooled murine antisera: one with homologous antiserum and two others with heterologous antisera, diluted 1:60,000 at TBST buffer containing 0.1% BSA. Membranes were then washed three times for 5 minutes with TBST buffer, incubated with HRP-conjugated goat anti-mouse Fc $\gamma$  diluted 1: 100,000 in TBST buffer with 0.1% BSA for 1 hour at room temperature and washed again three times for 5 minutes with TBST buffer. Chemiluminescent detection was carried out with Immobilon Western Chemiluminescent HRP Substrate and exposed to Kodak photographic film. Films were digitized with G:BOX imaging system (Syngene Inc.) The signal intensities of individual antigen reactions were compared and scored semi-quantitatively as described previously<sup>55</sup> with quartile modification. Briefly, the intensity of several spots could not be analyzed directly due to their overlapping and notable differences in size. Therefore, we firstly collected the equal sections of the primary images, containing whole spots or the middle parts of the spots (depending on whether the given spot was smaller or larger than the selected area). The procedure was repeated for any sections containing image background, in order to compare and eliminate differences in the overall saturation of the photographic films. Subsequently, the intensity of the spot samples was evaluated against the respective samples of the image background, using Image Studio Lite software (LiCor Inc.). Results were presented as percent values relative to the most intense spot.

**In gel trypsin digestion for LC MS/MS analysis.** Gels were rinsed with HPLC-grade water. Excised spots were destained in 100  $\mu$ l of 100 mM ammonium bicarbonate/acetonitrile solution for 30 minutes at room temperature and then washed with 500  $\mu$ l of neat acetonitrile. Gel pieces were then covered with 10 ng/ $\mu$ l porcine trypsin solution in 10 mM ammonium bicarbonate/10% (v/v) acetonitrile and incubated for 2 h on ice followed by overnight incubation at 37 °C. After digestion, samples were centrifuged and supernatant aliquots were withdrawn, and stored at –20 °C until LC MS/MS analysis.

**Identification of immunodominant membrane proteins by cross-reactive antibodies.** The gels were manually aligned on the top of blots by using selected OMP CD, OMP E and Cop B proteins as landmark proteins as they were clearly identified in Coomassie blue stained gels from both two strains. High sequence coverage in mass spectrometry for both the proteins was observed.

**Peptide identification by LC MS/MS analysis.** The analyses were performed on an Ion Trap LC/MS/MS spectrometer (Agilent Technologies, Santa Clara, CA USA). The resulting peptide mass fingerprints and LC MS/MS fragmentation spectra were identified using the MASCOT (<http://www.matrixscience.com>) and BLAST engines<sup>56</sup> searching *M. catarrhalis* protein databases.

**SDS-PAGE and Western blots.** For resolution of high molecular weight proteins present in OMVs, the 6% SDS-PAGE stained with GelCode Blue Stain Reagent (Thermo Scientific, USA). Proteins were transferred at 20 V overnight at 4 °C to an PVDF membrane (Millipore, USA) using Towbin buffer. The membranes were then blocked for 2 h with PBS supplemented with Tween 20 (T-PBS) and 3% BSA (SERVA). After 3 washes with T-PBS, the membrane strips were incubated with homologous or heterologous murine antisera in T-PBS-1% BSA for 2 h at 37 °C. After subsequent 3 washes, membrane strips were incubated with HRP-conjugated goat anti-mouse IgG Fc $\gamma$  diluted 1,000 in T-PBS-1% BSA for 2 h at 37 °C. Detection of proteins on blot transfer was carried out using substrates: 4-chloro-1-naphthol and perhydrol.

**Immunization and preparation of mouse antisera.** The antisera were obtained as previously described<sup>26</sup>. Briefly, groups of 5–7 BALB/c mice of approximately equal weight, at 8–10 weeks old were immunized intraperitoneally with heat inactivated *M. catarrhalis* in saline ( $5 \times 10^7$  cfu dose<sup>-1</sup>) in 0.1 ml doses, three times, at 2-week intervals (on days 0, 14 and 28). Ten days after the last immunization, mice were terminally bled through the retro-orbital sinus. Sera from each group immunized with particular immunogen (Mc1, Mc6 or Mc8) were pooled, divided and stored in aliquots at –70 °C until use. All experimental protocols were approved by the Ethical Review Committee of University of Environmental and Life Sciences of Wrocław, Poland. This study complied with the animal experimentation guidelines of the National Institutes of Health guide for the care and use of Laboratory animals (NIH Publications No. 8023, revised 1978).

**Quantitative and Avidity ELISA.** The relative avidities (AI) of antibodies to the OMV antigens, were determined by ELISA as described previously<sup>26</sup>.

**Genome sequencing.** Four *Moraxella* genomes were sequenced using the Illumina MiSeq platform (University of Northumbria, UK). A 2\*250 bp paired-end library (Nextera XT sample prep) were utilized for each sample. The raw data was processed by trimming the library adapter sequences and quality filtering was applied using CLC genomics workbench v.7.5 (CLC Bio, Aarhus, Denmark). Quality reads below 0.05, along with reads shorter than 100 bp and with 2 or more ambiguities were removed (total values removed for strain: MC1 – 0.06%; MC5 – 16.91%, MC6 – 7.55%, MC8 – 0.05%). After trimming and filtering, the data were processed and the reads were mapped to the reference genomes. The processed reads were used for *de novo* assembly. A high-quality draft genome was assembled by closing the gaps between the scaffolded contigs of the reference assembly by contig stitching with *de novo* contigs. The four final draft genomes comprised from 3 to 5 contigs were grouped in one scaffold each. The final genome assemblies were deposited in GenBank under accession numbers CP010573, CP010900, CP010901 and CP010902. Their detailed characteristics is provided in Supplementary Table 1.

The sequence of a pseudogene SQ02\_07885 in the Mc8 genome was verified by PCR analysis using the primers, F: gataagccgtataccgccattcatg, R: ctgtctgagtgttctctgcgccag. The obtained 1.2 kb product was resequenced (Genomed, Poland).

**Preparing sequence data for bioinformatics analysis.** Four newly sequenced and all publicly available 23 full genomes from *M. catarrhalis* strains as well as four other *Moraxella* species were included in the bioinformatics analysis (Supplementary Table 2). Genomes that were not fully sequenced, assembled or annotated were excluded because they would bias the presence and distribution of genes. As an outgroup three representatives of *Acinetobacter* were selected because they are close relatives to *Moraxella*, grouped together to family Moraxellaceae. The data for these genomes were downloaded from the GenBank database (<ftp.ncbi.nlm.nih.gov/genomes>). Their annotated protein sequences were clustered into orthologous groups using a genome-scale algorithm OrthoMCL<sup>57</sup> assuming BLAST cut-off for sequence similarity 25% and E-value threshold  $10^{-3}$ .

Phylogenetic relationships between the genomes studied were inferred using concatenated alignments of sequences from 739 orthologous groups that were represented by one sequence in each of all 34 genomes. The sequences were aligned with a slow and accurate algorithm PSI-Coffee, which combines homology extension and consistency based progressive alignment<sup>58</sup>. Sites appropriate for phylogenetic analyses were selected in Gblocks using less stringent criteria<sup>59</sup>. In the final analyses, the alignment with 235,616 sites was used.

**Searching for domains in protein sequences and their classification to functional categories.** The presence of protein domains in OMV proteins were analyzed by local searches of Conserved Domain Database, CDD<sup>60</sup> using rpsBLAST assuming E-value < 0.01. Protein sequences from the studied genomes were classified to more general functional categories according to Clusters of Orthologous Groups (COGs) being a part of the CDD database.

**Inferring phylogenetic relationships based on orthologous sequences.** In the construction of phylogenetic trees, a Bayesian approach was applied with MrBayes<sup>61</sup> and maximum likelihood method in more-PhyML<sup>62</sup> using PhyML<sup>63</sup>. In the MrBayes analysis, a mixed +  $\Gamma$  + I model of amino acid substitutions was used to specify appropriate models across the larger space in the Bayesian MCMC analysis<sup>64</sup>, avoiding the need for a



*priori* model testing. In this approach, two independent runs starting from random trees, each using eight Markov chains were applied. Trees were sampled every 100 generations for 1,000,000 generations. In the final analysis, trees from the last 703,000 generations that reached the stationary phase and convergence (i.e. the standard deviation of split frequencies stabilized and was lower than the proposed threshold of 0.01) were selected. The best-fit substitution model LG + F +  $\Gamma$  + I found among all 120 candidates in ProtTest<sup>65</sup> and the best search algorithm NNI + SPR were applied in morePhyML. To assess the significance of the branches, non-parametric bootstrap analysis was performed on 1000 replicates in PhyML.

The percentage difference between sequences of OMV proteins and the rest of orthologous groups, as well as homologous sequences of proteins and their epitope sites, analyzed in cross-reaction studies was calculated using ClustalW (Larkin *et al.* 2007) for alignments obtained in PSI-Coffee. To avoid a bias resulting from the presence and absence of proteins in different phylogenetic lineages, only sequences represented in all 34 genomes in the comparison of OMV and other orthologous protein groups were analysed.

**Inferring phylogenetic relationships based on presence/absence of genes.** Phylogenetic relationships were calculated based on the presence and absence of 4780 orthologous groups found by OrthoMCL in least two genomes. The presence of some genes only in 28 genomes were also included in the dataset. Two methods were applied, Bayesian in MrBayes<sup>61</sup> and maximum parsimony in PAUP<sup>66</sup>. The characters were considered as unordered. In MrBayes approach, the gamma-distributed rate model and two independent runs starting from random trees using 4 Markov chains were applied. Trees were sampled every 100 generations for 10,000,000 generations. In the final analysis, trees from the last 2,716,000 generations that reached the stationary phase and convergence were selected. In PAUP\*, the final tree was searched from 10 starting trees obtained by stepwise addition of characters' sequences followed by the tree-bisection-reconnection (TBR) branch-swapping algorithm. In the non-parametric bootstrap analyses, we assumed 1000 replicates.

**Inferring ancestral sizes of gene content.** Ancestral sizes of gene content in the genomes as well as gene gains and losses per phylogenetic lineage were determined in BadiRate<sup>67</sup> under the free rates birth-death-and-innovation model and maximum likelihood estimation procedure. Phylogenetic relationships were assumed as deduced in the phylogenetic analysis in MrBayes, whereas data about the size of homologous protein families in the studied genomes were taken from OrthoMCL results.

**Computational prediction of potential epitope sites.** Sequences of immunoreactive proteins were searched for potential epitopes using 15 predictive approaches: AAP<sup>68</sup>, BCPred<sup>69</sup>, BepiPred<sup>70</sup>, CBTOPE<sup>71</sup>, COBEpro<sup>72</sup>, FBCPred<sup>73</sup>, IgPred<sup>74</sup>, SVMTrip<sup>75</sup>, as well as BcePred<sup>76</sup> based on seven amino acid physico-chemical scales, hydrophilicity, flexibility, accessibility, turns, antigenic propensity, exposed residues and polarity. These results were summarized by counting the number of times individual sites in the sequences were predicted as an epitope<sup>77</sup>.

**Statistical analyses.** Statistical analyses were carried out in R package. Statistical significance between the numbers of protein sequences in functional categories was estimated using likelihood ratio (G-test) test of independence and test of proportions. P-values were corrected with Benjamini–Hochberg procedure for multiple testing. Differences in contribution of predicted epitope sites in sequences of proteins showing various cross-reactivity were analysed in t-test.

**Data Availability.** The genome sequences of Mc1, Mc5, Mc6 and Mc8 can be found in GenBank under accession numbers: CP010573, CP010900, CP010901, and CP010902 respectively. Most data generated or analyzed during this study are included in this published article (and its Supplementary Information files). Other data supporting this study are available from the corresponding author upon reasonable request.

## References

- Murphy, T. F., Brauer, A. L., Aebi, C. & Sethi, S. Antigenic specificity of the mucosal antibody response to *Moraxella catarrhalis* in chronic obstructive pulmonary disease. *Infect. Immun.* **73**, 8161–8166 (2005).
- Hays, J. P. *Moraxella catarrhalis*: a mini review. *J. Pediatric Infect. Dis.* **4**, 211–220 (2009).
- Liu, D. F., McMichael, J. C. & Baker, S. M. *Moraxella catarrhalis* outer membrane protein cd elicits antibodies that inhibit cd binding to human mucin and enhance pulmonary clearance of *M. catarrhalis* in a mouse model. *Infect. Immun.* **75**, 2818–2825 (2007).
- Smidt, M. *et al.* Comprehensive antigen screening identifies *Moraxella catarrhalis* proteins that induce protection in a mouse pulmonary clearance model. *PLoS ONE* **8**, e64422, <https://doi.org/10.1371/journal.pone.0064422> (2013).
- Ren, D. *et al.* Serum antibody response to *Moraxella catarrhalis* proteins OMP CD, OppA, Msp22, Hag, and PilA2 after nasopharyngeal colonization and acute otitis media in children. *Vaccine* **33**, 5809–5814 (2015).
- Ren, D. & Pichichero, M. E. Vaccine targets against *Moraxella catarrhalis*. *Expert Opin. Ther. Targets* **20**, 19–33 (2016).
- Bootsma, H. J., van der Heide, H. G., van de Pas, S., Schouls, L. M. & Mooi, F. R. Analysis of *Moraxella catarrhalis* by DNA typing: evidence for a distinct subpopulation associated with virulence traits. *J. Infect. Dis.* **181**, 1376–1387 (2000).
- Wirth, T. *et al.* The rise and spread of a new pathogen: seroresistant *Moraxella catarrhalis*. *Genome Res.* **17**, 1647–1656 (2007).
- Earl, J. P. *et al.* Comparative genomic analyses of the *Moraxella catarrhalis* serosensitive and seroresistant lineages demonstrate their independent evolution. *Genome Biol. Evol.* **8**, 955–974 (2016).
- Verhaegh, S. J. C. *et al.* Age-related genotypic and phenotypic differences in *Moraxella catarrhalis* isolates from children and adults presenting with respiratory disease in 2001–2002. *Microbiology* **154**, 1178–1184 (2008).
- Davie, J. J. *et al.* Comparative analysis and supragenome modeling of twelve *Moraxella catarrhalis* clinical isolates. *BMC Genomics* **12**, 1–17, <https://doi.org/10.1186/1471-2164-12-70> (2011).
- Blakeway, L. V. *et al.* ModM DNA methyltransferase methylome analysis reveals a potential role for *Moraxella catarrhalis* phasevarians in otitis media. *FASEB J* **28**, 5197–5207 (2014).
- de Vries, S. P. W., Bootsma, H. J., Hays, J. P. & Hermans, P. W. M. Molecular aspects of *Moraxella catarrhalis* pathogenesis. *Microbiol. Mol. Biol. Rev.* **73**, 389–406 (2009).
- Su, Y.-C., Singh, B. & Riesbeck, K. *Moraxella catarrhalis*: from interactions with the host immune system to vaccine development. *Future Microbiol.* **7**, 1073–1100 (2012).



15. Liu, G. *et al.* *Moraxella catarrhalis* evades host innate immunity via targeting cartilage oligomeric matrix protein. *J. Immunol.* **196**, 1249–1258 (2016).
16. Vidakovic, M. L. A. P. *et al.* B cell activation by outer membrane vesicles - a novel virulence mechanism. *PLoS Pathog.* **6**, 1–19, <https://doi.org/10.1371/journal.ppat.1000724> (2010).
17. Heiniger, N., Spaniol, V., Troller, R., Vischer, M. & Aebi, C. A reservoir of *Moraxella catarrhalis* in human pharyngeal lymphoid tissue. *J. Infect. Dis.* **196**, 1080–1087 (2007).
18. Pearson, M. M., Laurence, C. A., Guinn, S. E. & Hansen, E. J. Biofilm formation by *Moraxella catarrhalis* *in vitro*: roles of the UspA1 adhesin and the Hag hemagglutinin. *Infect. Immun.* **74**, 1588–1596 (2006).
19. Perez, A. C. *et al.* Residence of *Streptococcus pneumoniae* and *Moraxella catarrhalis* within polymicrobial biofilm promotes antibiotic resistance and bacterial persistence *in vivo*. *Pathog. Dis.* **70**, 280–288 (2014).
20. Parameswaran, G. I., Wrona, C. T., Murphy, T. F. & Sethi, S. *Moraxella catarrhalis* acquisition, airway inflammation and protease-antiprotease balance in chronic obstructive pulmonary disease. *BMC Infect. Dis.* **9**, 1–10, <https://doi.org/10.1186/1471-2334-9-178> (2009).
21. Kaparakis-Liaskos, M. & Ferrero, R. L. Immune modulation by bacterial outer membrane vesicles. *Nat. Rev. Immunol.* **15**, 1–13, <https://doi.org/10.1038/nri3837> (2015).
22. Tan, T. T., Mörgelin, M., Forsgren, A. & Riesbeck, K. *Haemophilus influenzae* survival during complement-mediated attacks is promoted by *Moraxella catarrhalis* outer membrane vesicles. *J. Infect. Dis.* **195**, 1661–1670 (2007).
23. Schaar, V., Nordström, T., Mörgelin, M. & Riesbeck, K. *Moraxella catarrhalis* outer membrane vesicles carry  $\beta$ -lactamase and promote survival of *Streptococcus pneumoniae* and *Haemophilus influenzae* by inactivating amoxicillin. *Antimicrob. Agents Chemother.* **55**, 3845–3853 (2011).
24. Augustyniak, D., Mleczko, J. & Gutowicz, J. The immunogenicity of the liposome-associated outer membrane proteins (OMPs) of *Moraxella catarrhalis*. *Cell Mol. Biol. Lett.* **15**, 70–89 (2010).
25. Augustyniak, D., Majkowska-Skrobek, G., Roszkowiak, J. & Dorotkiewicz-Jach, A. Defensive and offensive cross-reactive antibodies elicited by pathogens: The Good, the Bad and the Ugly. *Curr. Med. Chem.* **24**, 4002–4037 (2017).
26. Augustyniak, D., Piekut, M., Majkowska-Skrobek, G. & Skala, J. Bactericidal, opsonophagocytic and anti-adhesive effectiveness of cross-reactive antibodies against *Moraxella catarrhalis*. *Pathog. Dis.* **73**, ftu026 <https://doi.org/10.1093/femspd/ftu026> (2015).
27. Augustyniak, D. *et al.* Innate immune properties of selected human neuropeptides against *Moraxella catarrhalis* and nontypeable *Haemophilus influenzae*. *BMC Immunol.* **13**, 24, <https://doi.org/10.1186/1471-2172-13-24> (2012).
28. Schaar, V. *et al.* Multicomponent *Moraxella catarrhalis* outer membrane vesicles induce an inflammatory response and are internalized by human epithelial cells. *Cell. Microbiol.* **13**, 432–449 (2011).
29. Cope, L. D. *et al.* Characterization of the *Moraxella catarrhalis* uspA1 and uspA2 genes and their encoded products. *J. Bacteriol.* **181**, 4026–4034 (1999).
30. Blakeway, L. V., Tan, A., Peak, I. R. A. & Seib, K. L. Virulence determinants of *Moraxella catarrhalis*: distribution and considerations for vaccine development. *Microbiology (Reading, Engl.)* **163**, 1371–1384 (2017).
31. Vidakovic, M. L. & Riesbeck, K. Virulence mechanisms of *Moraxella* in the pathogenesis of infection. *Curr. Opin. Infect. Dis.* **22**, 279–285 (2009).
32. Hess, J. F. The *Moraxella bovis* RTX toxin locus mbx defines a pathogenicity island. *J. Med. Microbiol.* **55**, 443–449 (2006).
33. Angelos, J. A., Lane, V. M., Ball, L. M. & Hess, J. F. Recombinant *Moraxella bovoculi* cytotoxin-ISCAM matrix adjuvanted vaccine to prevent naturally occurring infectious bovine keratoconjunctivitis. *Vet. Res. Commun.* **34**, 229–239 (2010).
34. Schwechheimer, C. & Kuehn, M. J. Outer-membrane vesicles from Gram-negative bacteria: biogenesis and functions. *Nat. Rev. Microbiol.* **13**, 605–619 (2015).
35. Liu, Q. *et al.* Outer membrane vesicles from flagellin-deficient *Salmonella enterica* serovar Typhimurium induce cross-reactive immunity and provide cross-protection against heterologous *Salmonella* challenge. *Sci. Rep.* **6**, 1–13, <https://doi.org/10.1038/srep34776> (2016).
36. Nieves, W. *et al.* A *Burkholderia pseudomallei* outer membrane vesicle vaccine provides protection against lethal sepsis. *Clin. Vaccine Immunol.* **21**, 747–754 (2014).
37. Stålhammar-Carlén, M., Areschoug, T., Larsson, C. & Lindahl, G. Cross-protection between group A and group B streptococci due to cross-reacting surface proteins. *J. Infect. Dis.* **182**, 142–149 (2000).
38. Booy, R. *et al.* Cross-reacting antibodies against the pandemic (H1N1) 2009 influenza virus in older Australians. *Med. J. Aust.* **194**, 19–23 (2011).
39. Hallström, T. *et al.* *Pseudomonas aeruginosa* uses dihydrolipoamide dehydrogenase (Lpd) to bind to the human terminal pathway regulators vitronectin and clusterin to inhibit terminal pathway complement attack. *PLoS ONE* **10**, 1–21, <https://doi.org/10.1371/journal.pone.0137630> (2015).
40. Shinoy, M. *et al.* Immunoproteomic Analysis of proteins expressed by two related pathogens, *Burkholderia multivorans* and *Burkholderia cenocepacia*, during human infection. *PLoS ONE* **8**, e80796, <https://doi.org/10.1371/journal.pone.0080796> (2013).
41. Khan, M. N., Shukla, D., Bansal, A., Mustoori, S. & Ilavazhagan, G. Immunogenicity and protective efficacy of GroEL (hsp60) of *Streptococcus pneumoniae* against lethal infection in mice. *FEMS Immunol. Med. Microbiol.* **56**, 56–62 (2009).
42. Álvarez, A. *et al.* Association of the P64k dihydrolipoamide dehydrogenase to the *Neisseria meningitidis* membrane. *Biotechnol. Appl.* **21**, 137–142 (2004).
43. Hallström, T. *et al.* Dihydrolipoamide dehydrogenase of *Pseudomonas aeruginosa* is a surface-exposed immune evasion protein that binds three members of the factor h family and plasminogen. *J. Immunol.* **189**, 4939–4950 (2012).
44. Dennehy, R. & McClean, S. Immunoproteomics: the key to discovery of new vaccine antigens against bacterial respiratory infections. *Curr. Protein Pept. Sci.* **13**, 807–815 (2012).
45. Yu, R. H., Bonnah, R. A., Ainsworth, S. & Schryvers, A. B. Analysis of the immunological responses to transferrin and lactoferrin receptor proteins from *Moraxella catarrhalis*. *Infect. Immun.* **67**, 3793–3799 (1999).
46. West, D. *et al.* Recombinant *Neisseria meningitidis* transferrin binding protein A protects against experimental meningococcal infection. *Infect. Immun.* **69**, 1561–1567 (2001).
47. Furano, K. Identification of a conserved *Moraxella catarrhalis* haemoglobin-utilization protein, MhuA. *Microbiology* **151**, 1151–1158 (2005).
48. Khan, A. G. *et al.* High-affinity binding by the periplasmic iron-binding protein from *Haemophilus influenzae* is required for acquiring iron from transferrin. *Biochem. J.* **404**, 217–225 (2007).
49. Gómez, J. A., Criado, M. T. & Ferreiros, C. M. Bactericidal activity of antibodies elicited against the *Neisseria meningitidis* 37-kDa ferric binding protein (FbpA) with different adjuvants. *FEMS Immunol. Med. Microbiol.* **20**, 79–86 (1998).
50. Wedege, E. *et al.* Functional and specific antibody responses in adult volunteers in new Zealand who were given one of two different meningococcal serogroup B outer membrane vesicle vaccines. *Clin. Vaccine Immunol.* **14**, 830–838 (2007).
51. Sinha, N., Mohan, S., Lipschultz, C. A. & Smith-Gill, S. J. Differences in electrostatic properties at antibody-antigen binding sites: implications for specificity and cross-reactivity. *Biophys. J.* **83**, 2946–2968 (2002).
52. Peng, H.-P., Lee, K. H., Jian, J.-W. & Yang, A.-S. Origins of specificity and affinity in antibody-protein interactions. *Proc. Natl. Acad. Sci. USA* **111**, E2656–E2665, <https://doi.org/10.1073/pnas.1401131111> (2014).
53. Qiu, T. *et al.* Proteochemometric modeling of the antigen-antibody interaction: new fingerprints for antigen, antibody and epitope-paratope interaction. *PLoS ONE* **10**, e0122416, <https://doi.org/10.1371/journal.pone.0122416> (2015).

54. Bakri, F., Brauer, A. L., Sethi, S. & Murphy, T. F. Systemic and mucosal antibody response to *Moraxella catarrhalis* after exacerbations of chronic obstructive pulmonary disease. *J. Infect. Dis.* **185**, 632–640 (2002).
55. Williams, J. N., Weynants, V., Poolman, J. T., Heckels, J. E. & Christodoulides, M. Immuno-proteomic analysis of human immune responses to experimental *Neisseria meningitidis* outer membrane vesicle vaccines identifies potential cross-reactive antigens. *Vaccine* **32**, 1280–1286 (2014).
56. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
57. Li, L., Stoeckert, C. J. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
58. Di Tommaso, P. *et al.* T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* **39**, W13–W17, <https://doi.org/10.1093/nar/gkr245> (2011).
59. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577 (2007).
60. Marchler-Bauer, A. *et al.* CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res.* **41**, D348–52, <https://doi.org/10.1093/nar/gks1243> (2013).
61. Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
62. Criscuolo, A. morePhyML: Improving the phylogenetic tree space exploration with PhyML 3. *Mol. Phylogenet. Evol.* **61**, 944–948 (2011).
63. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
64. Huelsenbeck, J. P. Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Mol. Biol. Evol.* **21**, 1123–1133 (2004).
65. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164–1165 (2011).
66. Swofford, D. L. *PAUP: phylogenetic analysis using parsimony and other methods*, 4<sup>th</sup> ed. (Sinauer Associates Sunderland, MA, 1998).
67. Librado, P., Vieira, F. G. & Rozas, J. BadiRate: estimating family turnover rates by likelihood-based methods. *Bioinformatics* **28**, 279–281 (2012).
68. Chen, J., Liu, H., Yang, J. & Chou, K. C. Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. *Amino Acids* **33**, 423–428 (2007).
69. EL-Manzalawy, Y., Dobbs, D. & Honavar, V. Predicting linear B-cell epitopes using string kernels. *J. Mol. Recognit.* **21**, 243–255 (2008).
70. Larsen, J. E. P., Lund, O. & Nielsen, M. Improved method for predicting linear B-cell epitopes. *Immunome Res.* **2**, 1–7, <https://doi.org/10.1186/1745-7580-2-2> (2006).
71. Ansari, H. R. & Raghava, G. P. Identification of conformational B-cell Epitopes in an antigen from its primary sequence. *Immunome Res.* **6**, 6, <https://doi.org/10.1186/1745-7580-6-6> (2010).
72. Sweredoski, M. J. & Baldi, P. COBEpro: a novel system for predicting continuous B-cell epitopes. *Protein Eng. Des. Sel.* **22**, 113–120 (2009).
73. EL-Manzalawy, Y., Dobbs, D. & Honavar, V. Predicting flexible length linear B-cell epitopes. *Comput. Syst. Bioinformatics Conf.* **7**, 121–132 (2008).
74. Gupta, S., Ansari, H. R. & Gautam, A. Open Source Drug Discovery Consortium & Raghava, G. P. S. Identification of B-cell epitopes in an antigen for inducing specific class of antibodies. *Biol. Direct.* **8**, 1–15, <https://doi.org/10.1186/1745-6150-8-27> (2013).
75. Yao, B., Zhang, L., Liang, S. & Zhang, C. SVMTriP: a method to predict antigenic epitopes using support vector machine to integrate tri-peptide similarity and propensity. *PLoS ONE* **7**, e45152, <https://doi.org/10.1371/journal.pone.0045152> (2012).
76. Saha, S. & Raghava, G. P. S. BcePred: Prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. *Artificial Immune Systems, Lecture Notes in Computer Science* 3239, (197–204. Springer, Berlin Heidelberg, 2004).
77. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

## Acknowledgements

The authors acknowledge COST (European Cooperation in the field of Scientific and Technical Research) action BM1003, “Microbial cell surface determinants of virulence as targets for new therapeutics in Cystic Fibrosis.” Publication costs of this work were supported by Wroclaw Center of Biotechnology program “The Leading National Research Center (KNOW) for years 2014–2018”. This work was supported also by research grant of University of Wroclaw (1016/S/IGiM/T-20).

## Author Contributions

Conception and design: D.A. and P.M.; Development of methodology: D.A., P.M. and S.M.; Acquisition of data: R.S., D.A., B.R., J.R., S.M., D.S. and P.M.; Analysis and interpretation of data: D.A. and P.M. Writing and review of the manuscript: D.A. and P.M.; Revision: S.M. and Z.D.-K.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-23029-7>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.