

RESEARCH ARTICLE

# Bioinformatic Analysis of *Chlamydia trachomatis* Polymorphic Membrane Proteins PmpE, PmpF, PmpG and PmpH as Potential Vaccine Antigens

Alexandra Nunes<sup>1</sup>, João P. Gomes<sup>1</sup>, Karuna P. Karunakaran<sup>2</sup>, Robert C. Brunham<sup>2\*</sup>

**1** Bioinformatics Unit, Department of Infectious Diseases, National Institute of Health, Lisbon, Portugal, **2** Vaccine Research Laboratory, University of British Columbia Centre for Disease Control, Vancouver, Canada

\* [robert.brunham@bccdc.ca](mailto:robert.brunham@bccdc.ca)



**OPEN ACCESS**

**Citation:** Nunes A, Gomes JP, Karunakaran KP, Brunham RC (2015) Bioinformatic Analysis of *Chlamydia trachomatis* Polymorphic Membrane Proteins PmpE, PmpF, PmpG and PmpH as Potential Vaccine Antigens. PLoS ONE 10(7): e0131695. doi:10.1371/journal.pone.0131695

**Academic Editor:** David M. Ojcius, University of the Pacific, UNITED STATES

**Received:** April 23, 2015

**Accepted:** June 5, 2015

**Published:** July 1, 2015

**Copyright:** © 2015 Nunes et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** AN is a recipient of a post-doctoral fellowship (SFRH/BPD/75295/2010) from Fundação para a Ciência e a Tecnologia (FCT).

**Competing Interests:** The authors have declared that no competing interests exist.

## Abstract

*Chlamydia trachomatis* is the most important infectious cause of infertility in women with important implications in public health and for which a vaccine is urgently needed. Recent immunoproteomic vaccine studies found that four polymorphic membrane proteins (PmpE, PmpF, PmpG and PmpH) are immunodominant, recognized by various MHC class II haplotypes and protective in mouse models. In the present study, we aimed to evaluate genetic and protein features of Pmps (focusing on the N-terminal 600 amino acids where MHC class II epitopes were mapped) in order to understand antigen variation that may emerge following vaccine induced immune selection. We used several bioinformatics platforms to study: *i*) Pmps' phylogeny and genetic polymorphism; *ii*) the location and distribution of protein features (GGA(I, L)/FxxN motifs and cysteine residues) that may impact pathogen-host interactions and protein conformation; and *iii*) the existence of phase variation mechanisms that may impact Pmps' expression. We used a well-characterized collection of 53 fully-sequenced strains that represent the *C. trachomatis* serovars associated with the three disease groups: ocular (N=8), epithelial-genital (N=25) and lymphogranuloma venereum (LGV) (N=20). We observed that PmpF and PmpE are highly polymorphic between LGV and epithelial-genital strains, and also within populations of the latter. We also found heterogeneous representation among strains for GGA(I, L)/FxxN motifs and cysteine residues, suggesting possible alterations in adhesion properties, tissue specificity and immunogenicity. PmpG and, to a lesser extent, PmpH revealed low polymorphism and high conservation of protein features among the genital strains (including the LGV group). Uniquely among the four Pmps, *pmpG* has regulatory sequences suggestive of phase variation. In aggregate, the results suggest that PmpG may be the lead vaccine candidate because of sequence conservation but may need to be paired with another protective antigen (like PmpH) in order to prevent immune selection of phase variants.

## Introduction

*Chlamydia trachomatis* is an obligate intracellular human bacterial pathogen, comprised of 15 to 18 major serovars. Serovars A-C are responsible for ocular infections that result in trachoma [1]. Serovars D-K cause sexually transmitted diseases such as cervicitis and pelvic inflammatory disease (PID), and globally are an important infectious cause of infertility and ectopic pregnancy in women [2]. Serovars L1-L3 also enter through the ano-urogenital tract, but may disseminate via infection of macrophages to regional draining lymph nodes, causing lymphogranuloma venereum (LGV) [3]. *C. trachomatis* is the major bacterial cause of sexually transmitted infections (STIs), accounting for ~106 million of the 500 million new cases of STIs that occur each year worldwide [4]. In Europe, almost half of the estimated 47 million STI cases are due to *C. trachomatis* [4], while nearly 1.4 million infections are reported each year in the United States [5]. Although the estimated global economic burden is uncalculated, over \$516 million are spent annually in direct medical costs on genital chlamydial infections in the United States, making *C. trachomatis* the most costly infection among the nonviral STIs [6–7].

Despite screening and treatment public health programs to control *Chlamydia*, the incidence of *C. trachomatis* infection has increased [4, 8–10]. Thus, there is an urgent need for an efficacious vaccine that prevents acquisition and transmission of infection and the development of pelvic inflammatory disease sequelae. Cumulative studies in animal models and human infections [8, 11–23] have shown that systemic and mucosal CD4 T cell-mediated immunity is necessary for protection against *C. trachomatis* infection. Among antigen candidates that have been studied [11, 15, 24], members of the polymorphic membrane protein family (PmpA-I) have shown promise as vaccine components as they are dominant antigenic targets for cellular immune responses [25–29]. Four of the nine Pmps (PmpE, PmpF, PmpG and PmpH) have been identified via immunoproteomics as dominant T-cell antigens with multiple MHC class II binding peptides for both *C. trachomatis* and *C. muridarum* and observed to be protective in the murine genital tract infection model [25, 28–30]. The fact that each Pmp generates different peptides recognized by different MHC class II haplotypes confers on them the capability of immunizing outbred populations [25]. In the murine model, a PmpG epitope was found to persist on splenic antigen presenting cells for at least 6 months [26]. Tetramer staining also demonstrated PmpG as one of the quantitatively dominant antigens recognized by murine CD4 T cells [17].

Pmps are *Chlamydia*-specific outer membrane proteins whose precise functions remain unknown, but which have been implicated in pathogenesis and host cell adherence. As typical type V autotransporters [31–32], Pmps are capable of translocating to the bacterial surface their N-terminal Sec-dependent leader sequence (passenger domain), containing multiple short repetitive motifs (GGA(I, L, V) and FxxN) [33]. The proteins may also undergo complex infection-dependent post-translational proteolytic processing [34–37]. These proteins mediate *in vitro* chlamydial attachment to human epithelial and endothelial cells [34, 38]. Previous bioinformatics analyses suggested that six of the 9 *pmps* are under positive selection either driving bacterial adaptation to specific niches (ocular conjunctiva, epithelial-genitalia and lymph nodes), or during a strains' diversification within a particular niche [39]. Numerous mutations within bioinformatically predicted HLA class I and II T-cell epitopes for the N-terminal domain for one Pmp (PmpF) have already been identified [40]. Pmps were also found in tissue culture to be variably expressed at the chlamydial cell surface, where each protein seemed subjected to an independent high-frequency on/off switching at the inclusion level [41].

Therefore we aimed to bioinformatically assess the allelic and phase variation of these antigens, in particular for the four most promising vaccine candidates PmpE, PmpF, PmpG and PmpH, in order to understand if putative antigenic escape variants may emerge following

vaccination. We performed a detailed bioinformatic analysis using a well-characterized collection of 53 strains whose genomes have been fully sequenced [42–43] and which represent *C. trachomatis* serovars associated with the three canonical disease groups: ocular, epithelial-genital and LGV. By encompassing strains from temporally and geographically diverse sources around the world, this collection should capture *C. trachomatis* genetic diversity. Overall, we analyzed 25 strains from epithelial-genital serovars, 20 strains from LGV serovars, and eight strains from ocular serovars.

## Materials and Methods

### Genetic and phylogenetic analyses

In order to assess the genetic variability of PmpE, PmpF, PmpG and PmpH within *C. trachomatis*, 53 genome sequences representative of distinct or same-serovar strains were retrieved from GenBank and aligned using the progressiveMauve algorithm of Mauve 2.3.1 [44] with the default parameters and an initial match seed weight appropriate for 1MB genomes. Basically, Progressive Mauve performs a recursive anchor search and a full gapped anchored alignment of the genome sequences using a modified MUSCLE algorithm. For all *pmps*, individual alignments were extracted from the whole-genome alignment, and visually inspected with MEGA6 software (<http://www.megasoftware.net>) for further correction.

For each *pmp*, MEGA6 was also used to estimate the number of gene variable sites and to compute overall mean distances and matrices of pairwise comparisons at both nucleotide and protein level, based on the number of differences and *p*-distance value (that calibrates the obtained differences relative to the total number of sites under comparison) among the 53 strains, along with the respective standard error estimates (bootstrap = 1000). Evaluation of variable sites and mean genetic distances was also performed within and between the three disease groups. We also evaluated the impact of the genetic heterogeneity of each group of strains in the overall polymorphism of each *pmp*, through a sliding-window analysis using the DNA polymorphism tool of the DnaSP software, version 5 [45], with a window size and step size of 15.

Phylogenetic relationships among strains were inferred with MEGA 6, by using the Neighbor-Joining method [46] in conjunction with a bootstrap re-sampling strategy (1000 replicates), as previously described [47]. Evolutionary nucleotide distances were estimated with the Kimura 2-parameter (K2P) model [48] that takes into account transitional and transversional substitution rates, while assuming identical nucleotide frequencies and invariable substitution rates among sites [49]. At the protein level, evolutionary distances were computed based on the number of differences. Because of the different lengths of the sequences, the pairwise-deletion option was chosen to remove all sites containing missing data or alignment gaps from all distance estimations, only when the need arose and not prior to the analysis.

In order to identify specific regions that most contribute to the phylogenetic segregation of taxa within each Pmp, we evaluated the similarity both among the 53 strains and between disease groups. Briefly, SimPlot 3.5.1 software [50] was used to plot a codon-based nucleotide similarity score of each strain against a particular query by estimating pairwise distances with K2P model, without excluding gaps among sequences and considering a transition-transversion substitution rate of 2. The similarity estimations were performed in a sliding window size that ranged from 40 bp for *pmpG* to 80 bp for *pmpF* (adjusted according to polymorphism degree), moved across the alignment in a step size of 3 bp. In parallel, SWAAP 1.0.3 software [51] was used to compute the percentage of amino acid (a.a) identity among each pair of sequences throughout all Pmps, over a sliding window that ranged from 20 a.a for both PmpG and PmpH to 40 a.a for PmpF, and a step size of 3 a.a.

These analyses were also performed using gene segments encoding the first ~600 amino acids of each Pmp, as they are part of the surface exposed passenger domain where all MHC class II binding peptides were experimentally mapped [25] (see [Results and Discussion](#) section).

## Analysis of protein features

Pmps are predicted to have adhesin functions [34, 38, 52]. By using the SeqBuilder module of LaserGene (DNASTAR) we therefore assessed the differential presence among all 53 strains of the repetitive GGA(I, L, V) and FxxN motifs as well as of cysteine residues in each Pmp. Moreover, to shed light on the putative impact of Pmps' variability on surface probability and immunogenicity, we used the Protean program (DNASTAR) to perform a comparative analysis of the protein sequences. While the surface probability parameter predicts the likelihood of a given region lying on the surface of a protein using the approach of Emini *et al.* [53], the analysis of immunogenicity uses the approach of Jameson and Wolf [54] that combines several methods for protein structural features (like hydrophilicity, surface probability, flexibility and secondary structure) to predict potential antigenic determinants. For both analyses, the default parameters of each method were used, with surface regions predicted by forming the product of residue specific surface propensities over a range of 5 amino acids and based on a surface decision threshold  $>1$ . In order to facilitate the latter analyses, we used one strain representative of the main branches of the respective tree generated with the 53 fully-sequenced genomes, as these strains accurately represent the Pmp genetic backbone of the remainder same-branch taxa. The ocular strain A/Har13, the two epithelial-genital strains D/UW3 and E/150, and the LGV strain L2b/UCH1 were selected as they were among the strains found to be commonly present in the different main branches of all phylogenetic trees. Considering that some Pmps have an additional divergent branch containing genital strains, we opted for including an extra epithelial-genital strain to increase the confidence of the analysis. Thus, the strain G/11074 was also used for the analysis of both PmpH and PmpG, while the strains G/11222 and E/SW2 were used for the analysis of PmpF and PmpE, respectively. After performing preliminary data, we observed that including other strains besides these ones, did not alter the final output. In order to increase the probability of observing relevant disparities among strains for surface probability and immunogenicity, we specially focused on regions falling below a similarity cut-off of 85% (based on SWAAP plots).

## Analysis of phase variation

Analysis of phase variation was performed by using different approaches by considering several genetic features known to underlie phase variation mechanisms, such as short sequence repeats (homopolymeric and non-homopolymeric tracts), IS-like sequences, frameshift mutations, small indels, hairpin structures, RNase E cleavage sites, and promoter sequences [55–57]. First, we checked the differential presence of homopolymeric tracts and small indels among all 53 strains by using SeqBuilder (DNASTAR). For each operon, *in silico* promoter predictions were made by using both the Neural Network Promoter Prediction (NNPP, [http://www.fruitfly.org/seq\\_tools/promoter.html](http://www.fruitfly.org/seq_tools/promoter.html)) and the BPROM software (Softberry, <http://linux1.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfindb>), to find elements resembling putative  $\sigma^{66}$ ,  $\sigma^{28}$  and/or  $\sigma^{54}$  promoter sequences that may be differentially present among strains, and thus used as fine-tune transcriptional *pmp* regulators. Moreover, several putative regulatory elements that are known to affect transcription or translation were also searched throughout each *pmp* operon and associated regulatory regions using SeqBuilder (DNASTAR). Briefly, the differential existence of putative consensus cleavage sites for RNase E was examined

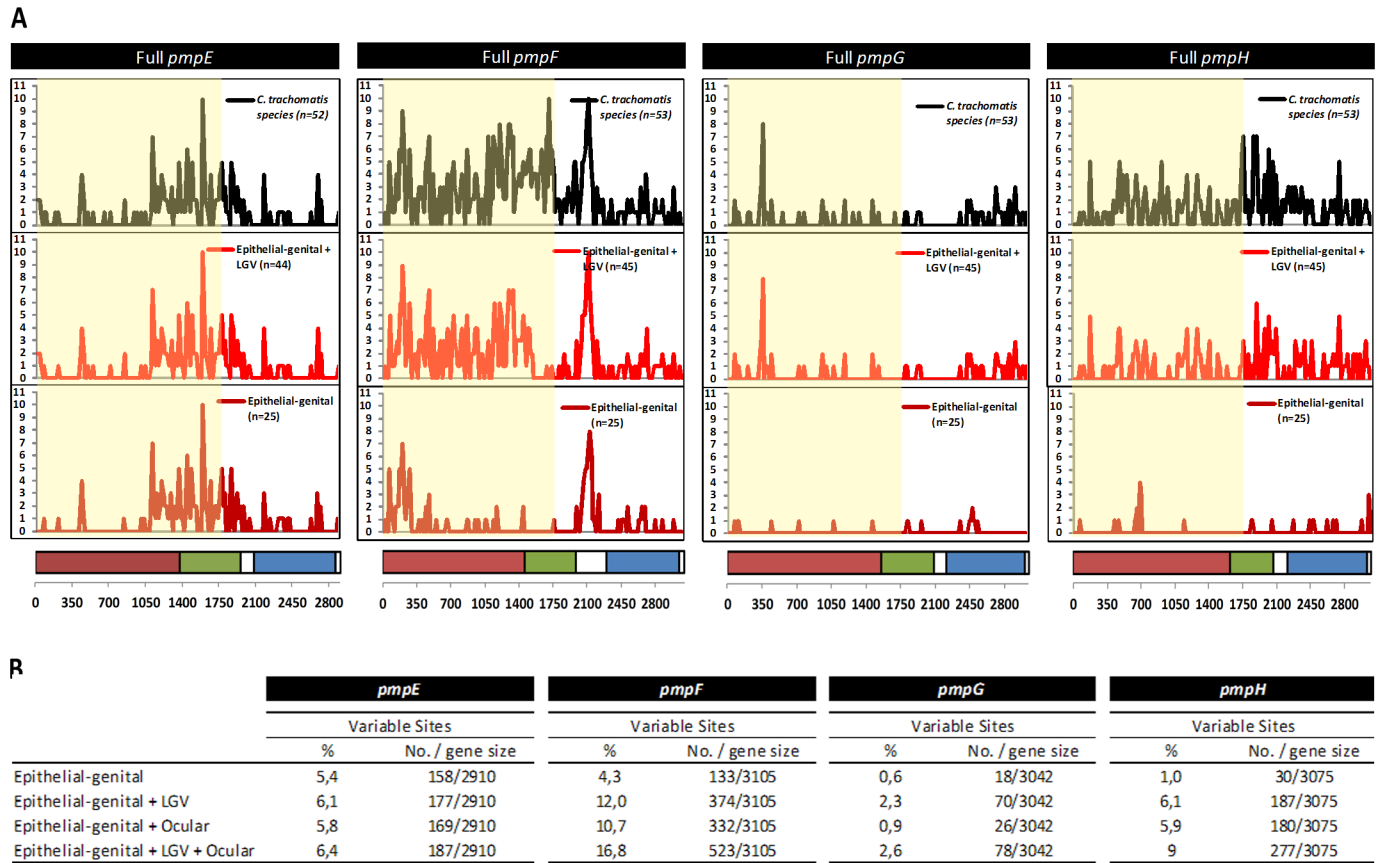
among strains [58–60]. RNaseE is the major endonuclease that generally initiates mRNA degradation in most bacteria [61]. We also searched for the presence of putative Shine-Dalgarno ribosome binding sequences (RBS) [62] and previously described chlamydial RBS [63–66].

## Results and Discussion

### Global polymorphism analysis

In view of *Chlamydiae*'s reductive evolution, it is remarkable that the *pmps* vary in number across species and encompass a sizeable chromosome portion (3–5%) [67]. For instance, over 13% of the *C. trachomatis* specific coding capacity is restricted to this gene family [68]. The evaluation of the mean genetic distances among all 53 strains revealed that *pmpF* was the most polymorphic *pmp* gene exhibiting the highest mean nucleotide substitution (214.3 (SE 8.9)) and also the highest *p*-distance value (0.0691 (SE 0.0029)), which corresponded to a mean of 72.4 (SE 5.0) (7.0%) amino acid substitutions. This degree of polymorphism is followed by *pmpH* and *pmpE* with an overall nucleotide variability of 3.6% and 2.4%, respectively. *pmpG* was observed to be the least variable gene with a polymorphism 6.9-fold lower than that of *pmpF*. In fact, *pmpG* exhibited a mean of nucleotide substitutions among all 53 strains of 31.1 (SE 3.7) (1.0%), corresponding to a mean of 13.5 (SE 2.6) (1.3%) amino acid alterations. The high variability shown by some Pmps accords with a recent analysis encompassing ~98% of the *C. trachomatis* core genome [69] which placed *pmpF*, *pmpE* and *pmpH* within the top 20 most polymorphic chromosomal genes. The set of polymorphic genes includes other important anti-gens-coding genes, such as CT681/*ompA* and the paralogously related *pmps* CT049-CT051 [70–72]. Considering Pmps' outer membrane localization and putative dual function in adhesion and pathogenesis, the high polymorphism is expected to promote multiple antigenic and/or adherence phenotypes that may influence strains' pathogenic diversity and tissue specificity.

By performing a polymorphism sliding-window analysis throughout each gene (Fig 1), we observed that most of the genetic variability in *pmpE* is concentrated into the mid-region of the gene (~1050–2100 bp), while a greater nucleotide substitution density is seen within the first two-thirds of *pmpF*. In contrast, for both *pmpG* and *pmpH*, the overall variability appears to be homogeneously distributed throughout each gene, although a polymorphism peak is seen around position 350 for *pmpG* (corresponding to three amino acids that are specific of LGV strains) We also found distinct scenarios among the four *pmps* by evaluating the contribution of the different disease groups to the overall species polymorphism (Fig 1). For both *pmpF* and *pmpH*, the three disease groups seem to contribute to *C. trachomatis* diversity, as seen by the successive increase in the overall genetic variability after the independent addition of either LGV or ocular strains to the group of epithelial-genital strains. In fact, the number of variant sites found among epithelial-genital strains augmented ~3-fold for *pmpF* and ~6-fold for *pmpH* with addition of LGV strains, and an increase of 1.5-fold was seen after the subsequent inclusion of ocular strains for both genes. For *pmpE*, the overall species polymorphism is essentially provided by the genetic heterogeneity among the epithelial-genital strains, which harbor almost 85% of the total 187 variant sites found. The dissimilar mutational pattern observed for each *pmp* may imply differences in how each Pmp fulfills adhesive and antigenic functions in specific niches. Distinct regions were previously identified within *pmpE*, *pmpF* and *pmpH*, where clusters of mutations were found, and were associated with strains' clustering by cell-tropism or ecological success [73]. In support of this notion, these four *pmps* were found to be differently targeted by positive selection in each niche, being potentially involved in one or more adaptive processes [39]. For instance, based on bioinformatic predictions that rely essentially on the distribution and exclusive character of nonsilent changes, *pmpF* is identified to encode specific cell tropism to both the ocular epithelium and mononuclear phagocytes.



**Fig 1. Global polymorphism analysis.** A) Sliding-window analysis of the genetic variability throughout each *pmp* (window and step size of 15 bp). For each *pmp*, the black plot represents the polymorphism among all strain collection, while the remainder plots represent the impact of removing strains from the ocular group (red plot) and from the LGV group (magenta plot) on the global polymorphism. The region highlighted in yellow encompasses the 1<sup>st</sup> 600 amino acids (used in vaccine attempts) where further analyses were performed. The horizontal bars below the plots represent the typical domains of these autotransporter proteins: passenger domain (magenta), middle domain (green), and C-terminal autotransporter domain (blue). B) Impact in both the number and percentage of variable sites found among *C. trachomatis* strains after the successive addition of LGV and ocular strains to the epithelial-genital group.

doi:10.1371/journal.pone.0131695.g001

Besides niche-specific adaptation, some of the *pmps* were also predicted to evolutionarily impact strains' niche-specific pathogenicity, as the case of *pmpG* and *pmpH*, which appear to contribute to the pathogenic diversity among LGV-causing strains [39]. Both *pmpG* and *pmpH* have over-accumulated nonsynonymous substitutions (about 4-fold more frequent than silent mutations) that result in protein variation among LGV strains [74]. Multiple factors may contribute to the genetic diversity of *pmps*, such as recombinational hotspots involving this large family of paralogous genes since multiple recombination hotspots have been identified throughout the genome [43, 75]. Overall, these data suggest the existence of variant tissue-specific host-interaction motifs that involve different Pmps.

### Genetic and phylogenetic analyses of Pmps' N-terminal 600 amino acid domains

Previous immunoproteomic studies using the murine genital infection model [25] revealed that PmpE, PmpF, PmpG and PmpH contain different MHC class II binding peptides in the N-terminal half for both *C. muridarum* and *C. trachomatis* serovar D with between one and three different peptides *per* MHC allele. Therefore, after an initial genetic analysis of the overall

**Table 1. Mean amino acid distances within the 1<sup>st</sup> 600 amino acids of each Pmp.**

	PmpE		PmpF		PmpG		PmpH	
	No. diff.	SE	No. diff.	SE	No. diff.	SE	No. diff.	SE
Overall mean	28,0	3,4	53,2	4,4	7,7	1,8	20,9	2,7
<i>Within Groups</i>								
Ocular	1,9	0,8	1,0	0,5	0,8	0,4	0,0	0,0
Epithelial-genital	14,5	2,0	14,0	2,3	1,7	0,9	2,0	0,8
LGV	0,0	0,0	0,0	0,0	0,6	0,3	0,8	0,4
Genital (with LGV)	19,5	2,5	52,2	4,7	7,4	1,8	16,9	2,7
<i>Between Groups</i>								
Ocular / Epithelial-genital	39,2	5,3	65,5	6,9	3,4	1,4	29,9	5,0
Ocular / LGV	47,4	6,3	111,0	8,9	11,6	3,1	45,5	6,3
Epithelial-genital / LGV	19,0	3,1	84,0	7,8	12,0	3,2	31,0	5,0

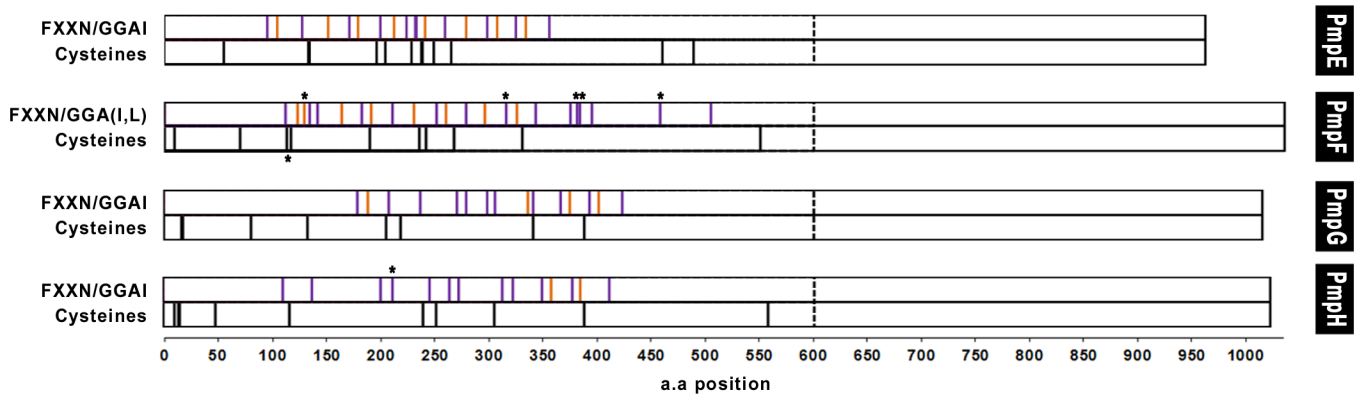
doi:10.1371/journal.pone.0131695.t001

polymorphism for the entire genes, the subsequent analyses (genetic, phylogenetic, and study of protein features) were restricted to gene segments encoding the 1<sup>st</sup> ~600 amino acids of each Pmp that encompass both the surface exposed passenger and middle domains (Fig 1) where all MHC class II-bound peptides were experimentally mapped [25]. The selected gene segments harbor almost 75% of all variant sites estimated for the whole *pmpF* and *pmpE* genes and at least 50% of the variant sites for *pmpG* and *pmpH*.

The phylogenetic reconstructions based on the first 600 amino acids of each Pmp (S1 Fig), mirror the tree topologies for the whole proteins (S2 Fig) and support the correlation between *pmp* polymorphism and a strains' phenotypic association with disease state. For instance, a perfect segregation of strains by full-tropism was observed for PmpH, with strains representative of the three disease groups appearing clustered in distinct separated clades. For all Pmps, the analysis of the matrices of pairwise distances revealed a low heterogeneity within both ocular and LGV strains (Table 1). In contrast, the average evolutionary divergence among the epithelial-genital group varies among Pmps, with PmpE and PmpF exhibiting the highest divergence, 14.5 a.a (SE 2.0) and 14.0 (SE 2.3), respectively, while PmpG and PmpH are much less polymorphic, displaying a polymorphism 7- to 9-fold lower. However, if LGV strains are included in a global group of genital strains, the average amino acid distances increases as epithelial-genital strains are highly dissimilar from LGV strains for some Pmps (Table 1). Overall, these results show that the 1<sup>st</sup> 600 a.a of PmpG is highly conserved in this 53 strain collection, even among all genital strains (epithelial-genital and LGV), for which the observed mean amino acid distances are about 2-, 3- and 7-fold lower than those seen for PmpH, PmpE and PmpF, respectively. Of note, the segment of the 1<sup>st</sup> 600 a.a of PmpH is also highly conserved if one considers solely epithelial-genital strains, where no more than 2 a.a differences are seen.

### Analysis of peptide features

A hallmark of the chlamydial Pmp family is the presence of multiple repeats of the tetrapeptide motifs GGA(I, L, V) and FxxN in their N-terminal half. The repetition of these motifs is seen in very few non-chlamydial proteins, and has been suggested to be directly involved in adherence processes [33, 76]. Thus, considering that Pmps likely have adhesive functions [34, 38] and were found to be variably expressed at the chlamydial surface *in vitro* [41], we mapped the location and determined the number of the GGA(I, L, V) and FxxN as well as of the cysteine residues within the 1<sup>st</sup> 600 a.a of each Pmp for all 53 strains (Fig 2). We observed that the FxxN motif occurs on average 12.3-times *per* Pmp, being ~2-fold more frequent than GGAI motifs,



**Fig 2. Distribution of FxxN motifs (purple), GGA(I, L) motifs (orange), and cysteine residues (black) within the 1<sup>st</sup> 600 a.a (limited by the vertical dashed line) of each Pmp.** Asterisks indicate motifs nonconserved among all 53 strains.

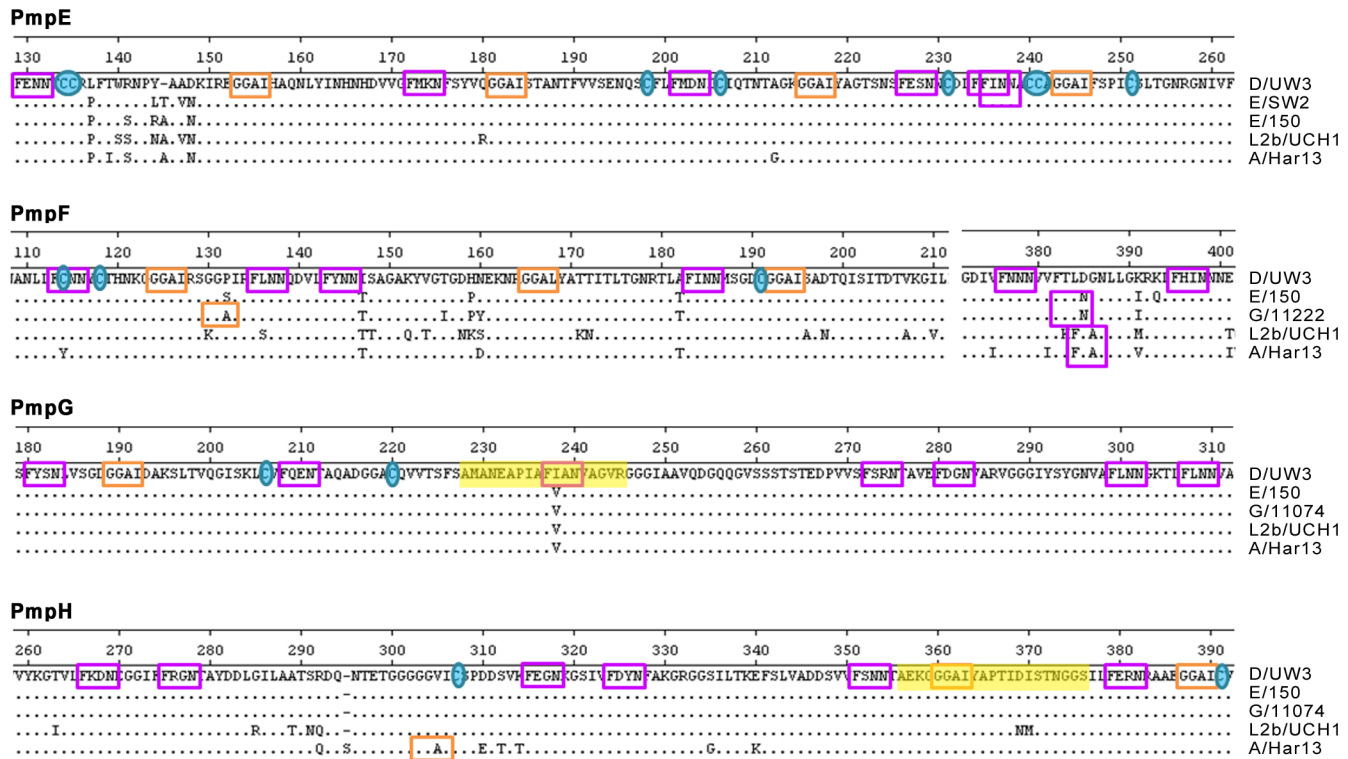
doi:10.1371/journal.pone.0131695.g002

which appear on average 5.8-times *per* Pmp. The mean incidence of FxxN in the remaining *C. trachomatis* proteome was shown to be 0.73 *per* protein, whereas GGAI is predicted to be present as single copy in only 10 other *Chlamydia* proteins [33]. No GGA(I, L, V) or FxxN motif is found outside the 1<sup>st</sup> 600 a.a for each Pmp, which is consistent with the exclusive presence of adhesion domains in protein regions that interact with the host. The number of conserved GGAI motifs varied among the four Pmps, ranging from two for PmpH to eight for PmpE. All Pmps displayed 11 conserved FxxN motifs; one of these fell within the MHC class II (I-A<sup>b</sup>)-bound *C. trachomatis* serovar D-derived peptide (AMANEAPIAFIANVAG) (Fig 3) recently identified in PmpG [25].

The distribution of the two types of peptide motifs is nonrandom, being mainly clustered within a discrete region that spans between ~220 a.a for PmpH to ~280 a.a for PmpF (Fig 2). Most of these motifs appear as doublets of FxxN-x<sub>2-25</sub>-FxxN and FxxN-x<sub>4-20</sub>-GGAI, whose differential occurrence, number and spacing vary both among Pmps and among different groups of strains for the same Pmp. They were also found to be in close proximity or adjacent to cysteine residues. The number of conserved cysteines ranges from eight for PmpG to 12 for PmpE. Considering that cysteines are known to play an important role in maintaining structural motifs, such GGAI/FxxN nonrandom distribution and possible cooperative action with the cysteine residues may yield unique structural or functional constraints for interaction with the host cell.

We also observed that the majority of the tetrapeptide motifs occurred in close proximity or fall within regions containing SNPs differentiating specific groups of strains (Fig 3), which may widen adhesion or immunogenic differences within each Pmp, as described for the major chlamydial membrane protein MOMP [77–79]. In support of this, a typical conserved FxxN-X<sub>5</sub>-GGAI doublet was found to overlap the MHC class II (I-A<sup>b</sup>)-bound peptide (AEKGG-GAIYAPTIDISTNGGS) identified in PmpH [25], which is conserved solely among ocular and epithelial-genital strains (Fig 3). This epitope is also in very close proximity to another conserved FxxN-X<sub>4</sub>-GGAI doublet and a conserved cysteine residue. Based on bioinformatic predictions, none of the three MHC class II (I-Ab)-bound *C. trachomatis* serovar D-derived peptides recently identified in PmpE, PmpG and PmpH [25] fall in regions that impact variability in immunogenicity and surface exposure of the respective proteins (S3 Fig). Although the precise function of the tetrapeptide motifs is yet unknown, it has been shown by using deletion analysis that Pmps’ adhesion capability requires at least the presence of one of these doublets [80]. Also, each Pmp was shown to exhibit *in vitro* a distinctive adhesion profile





**Fig 3. Pmp alignments showing examples of protein regions containing high concentration of FxxN motifs (purple), GGA(L, L) motifs (orange), and cysteine residues (blue). MHC class II (I-Ab)-bound *C. trachomatis* peptides are highlighted in yellow.**

doi:10.1371/journal.pone.0131695.g003

depending on the human cell type (epithelial *versus* endothelial) [34]. The mutational pattern of PmpE, PmpF, PmpG and PmpH was shown to be associated with high efficiency of *in vitro* attachment to host cells [81]. It has been hypothesized that these short repetitive motifs may be involved in maintaining the Pmp conformation that promotes adhesion, and/or might directly mediate interaction with human receptors [80]. Therefore, the wide variation of these repetitive motifs seen among intra- and inter-Pmps (Figs 2 and 3), is consistent with a role in the formation of niche-specific binding “receptors” for host-interaction.

### Analysis of phase variation

Since *pmps* have been suggested to undergo phase variation-like mechanisms to promote multiple antigenic and/or adherence phenotypes [41], we searched for putative phase variation features on each gene [55–57]. In distinction to *pmpE*, *pmpF* and *pmpH* that reveal no poly(C) or poly(G) stretches, *pmpG* exhibits an in-frame poly(G) tract of nine residues for all non-LGV strains and eight residues for LGV strains (data not shown). This mirrors that found for *C. pneumoniae pmpG* family [33, 82–84], suggesting that slipped-strand phase variation [55, 57] may be a common phenotype for PmpG. However, it is not known if the observed poly(G) tract influences *pmpG* expression in *C. trachomatis*. Although all annotated genome sequences appear to possess a functional PmpG, we cannot be certain that *C. trachomatis* strains only harbor in-frame poly(G) tract in *pmpG*. In fact, it is known that this type of regions may yield biased data in some Next Generation Sequencing. So, it is possible that strains with dissimilar number of “G” (or even a mixture of clones showing different number of “G”) have not been properly annotated. Nevertheless, as only PmpG shows a poly-G tract, which may be involved

in phase variation, and because its sequence is highly conserved among all genital strains (including LGV strains), this may be its major mechanism of antigen or niche variation. We also found a non-homopolymeric repetitive motif “PAPAPAPA” in PmpH for some epithelial-genital strains, which is shorter for other epithelial-genital strains and absent for both the LGV and ocular groups (data not shown). A Blast search for this motif was not informative about its putative role, as it seems to be commonly present in diverse proteins, such as regulatory proteins, acid-shock proteins and acetyl-CoA carboxylases, from different species. Hypothetically this genetic feature could be associated with phase variation and would exclusively impact PmpH immunogenicity of epithelial-genital strains. Although this hypothesis is speculative, if PmpH is used as a vaccine component, the use of peptides encompassing all “PA” combinations may be prudent.

It is known that the insertion-excision of mobile elements (like IS-elements) may also impact gene expression [55], and the presence of putative remnants of IS-like elements flanked by direct target repeats has already been described for both *pmpB* and *pmpC* [73]. However, no such elements were found in the four *pmps* under evaluation.

Past studies have shown that expression regulation of *C. trachomatis* genes is controlled at both transcriptional and translational levels, thus involving multiple complex aspects, like DNA supercoiling, heterogeneity within promoter sequences, cis- and trans-regulatory elements, and mRNA stability. It is known that various chlamydial genes are regulated by two or three different promoters, often with multiple  $\sigma$  factor binding ( $\sigma^{66}$ ,  $\sigma^{28}$  and/or  $\sigma^{54}$ ) [85]. We found putative  $\sigma^{66}$  promoters within both *pmpFE* and *pmpGH* regulatory regions by using *in silico* predictions (S1 and S2 Tables), but no elements resembling a  $\sigma^{28}$  or  $\sigma^{54}$  promoter sequence. In distinction to *pmpFE*, the most promoter predicted for the *pmpGH* operon contains a transcriptional start site (TSS) that has already been experimentally identified for the L2b/UCH1 strain [86], and it reveals a variable site in the -35 element that distinguishes LGV from epithelial-genital strains. Some variable nucleotide sites are also seen near the -10 elements, the TSSs, or A/T spacer, and several putative consensus RNase E cleavage sites were found throughout both operons and their regulatory regions. The polymorphisms found close or inside Shine-Dalgarno regulatory elements and the nonconservation of RNase E cleavage sites, may suggest heterogeneous expression among strains for both *pmpFE* and *pmpGH* operons, supporting previous experimental findings showing Pmp expression differences not only between L2 and E strains, but also within same-serovar strains [87].

## Conclusions

Recent vaccine studies in the murine model [29, 88] provided evidence that a vaccine composed of PmpEFGH plus MOMP formulated with a Th1 polarizing adjuvant, was more immunogenic and cleared infection faster than a single antigen vaccine. The suitability of a Pmp-based vaccine will be influenced by antigenic variability displayed during infection of the genital tract. The *in silico* predictions from the present study suggest that PmpF and PmpE may be less reliable antigens for vaccine purposes due to high sequence polymorphism. This polymorphism, together with putative alterations in structural constraints provided by heterogeneity among strains in GGA(I, L)/FxxN motifs and cysteine residues, also suggest a possible role in immunogenicity variability. By contrast, the low polymorphism and high conservation of protein features for PmpG and to a lesser extent for PmpH, suggest that these proteins may be better vaccine candidates. Since phase variation may impact PmpG expression a single-component PmpG subunit vaccine may not provide protection against infection due to phase variation. Based on bioinformatics analysis we suggest that pairing of PmpG with PmpH could be a viable approach, in order to provide a range of epitopes for CD4+ T cell recognition

among different MHC genetic backgrounds and to provide cross-protection against multiple antigenic variants of *C. trachomatis*.

## Supporting Information

**S1 Fig. Phylogenetic trees of the 1st 600 amino acids of each Pmp.**

(PDF)

**S2 Fig. Phylogenetic trees of each Pmp.**

(PDF)

**S3 Fig. Amino acid (a.a) identity throughout the 1st 600 a.a of PmpE (panel A), PmpF (panel B), PmpG (panel C) and PmpH (panel D) among strains representative of the main branches of each phylogenetic tree.**

(PDF)

**S1 Table. *In silico* promoter predictions for *pmpFE* operon regulatory region.**

(PDF)

**S2 Table. *In silico* promoter predictions for *pmpGH* operon regulatory region.**

(PDF)

## Acknowledgments

AN is a recipient of a post-doctoral fellowship (SFRH/BPD/75295/2010) from Fundação para a Ciência e a Tecnologia (FCT).

## Author Contributions

Conceived and designed the experiments: AN JPG RCB. Performed the experiments: AN JPG. Analyzed the data: AN JPG RCB. Contributed reagents/materials/analysis tools: AN JPG RCB. Wrote the paper: AN JPG KPK RCB.

## References

1. Vogel G. Infectious diseases. Tackling neglected diseases could offer more bang for the buck. *Science*. 2006; 311(5761):592–593. doi: 10.1126/science.311.5761.592a [pii] doi: [10.1126/science.311.5761.592a](https://doi.org/10.1126/science.311.5761.592a) PMID: [16456044](https://pubmed.ncbi.nlm.nih.gov/16456044/).
2. Peipert JF. Clinical practice. Genital chlamydial infections. *N Engl J Med*. 2003; 349(25):2424–2430. doi: [10.1056/NEJMc030542](https://doi.org/10.1056/NEJMc030542) 349/25/2424 [pii]. PMID: [14681509](https://pubmed.ncbi.nlm.nih.gov/14681509/).
3. Schachter J. Chlamydial infections (first of three parts). *N Engl J Med*. 1978; 298(8):428–435. PMID: [340951](https://pubmed.ncbi.nlm.nih.gov/340951/).
4. WHO. Global incidence and prevalence of selected curable sexually transmitted infections—2008. Geneva, Switzerland: 2012.
5. CDC. Sexually Transmitted Disease Surveillance 2013. Atlanta, GA: 2014.
6. Owusu-Edusei K Jr., Chesson HW, Gift TL, Tao G, Mahajan R, Ocfemia MC, et al. The estimated direct medical cost of selected sexually transmitted infections in the United States, 2008. *Sex Transm Dis*. 2013; 40(3):197–201. doi: [10.1097/OLQ.0b013e318285c6d2](https://doi.org/10.1097/OLQ.0b013e318285c6d2) 00007435-201303000-00003 [pii]. PMID: [23403600](https://pubmed.ncbi.nlm.nih.gov/23403600/).
7. Pultorak E, Wong W, Rabins C, Mehta SD. Economic burden of sexually transmitted infections: incidence and direct medical cost of Chlamydia, gonorrhoea, and syphilis among Illinois adolescents and young adults, 2005–2006. *Sex Transm Dis*. 2009; 36(10):629–636. doi: [10.1097/OLQ.0b013e3181a96d23](https://doi.org/10.1097/OLQ.0b013e3181a96d23) 00007435-200910000-00007 [pii]. PMID: [19955874](https://pubmed.ncbi.nlm.nih.gov/19955874/).
8. Brunham RC, Rappuoli R. Chlamydia trachomatis control requires a vaccine. *Vaccine*. 2013; 31(15): 1892–1897. doi: S0264-410X(13)00081-9 [pii] doi: [10.1016/j.vaccine.2013.01.024](https://doi.org/10.1016/j.vaccine.2013.01.024) PMID: [23375977](https://pubmed.ncbi.nlm.nih.gov/23375977/); pmcid: PMC4148049.

9. Gottlieb SL, Low N, Newman LM, Bolan G, Kamb M, Broutet N. Toward global prevention of sexually transmitted infections (STIs): the need for STI vaccines. *Vaccine*. 2014; 32(14):1527–1535. doi: S0264-410X(13)01398-4 [pii] doi: [10.1016/j.vaccine.2013.07.087](https://doi.org/10.1016/j.vaccine.2013.07.087) PMID: [24581979](https://pubmed.ncbi.nlm.nih.gov/24581979/).
10. Rekart ML, Brunham RC. Epidemiology of chlamydial infection: are we losing ground? *Sex Transm Infect*. 2008; 84(2):87–91. doi: sti.2007.027938 [pii] doi: [10.1136/sti.2007.027938](https://doi.org/10.1136/sti.2007.027938) PMID: [18216155](https://pubmed.ncbi.nlm.nih.gov/18216155/).
11. Brunham RC, Rey-Ladino J. Immunology of Chlamydia infection: implications for a Chlamydia trachomatis vaccine. *Nat Rev Immunol*. 2005; 5(2):149–161. doi: nri1551 [pii] doi: [10.1038/nri1551](https://doi.org/10.1038/nri1551) PMID: [15688042](https://pubmed.ncbi.nlm.nih.gov/15688042/).
12. Farris CM, Morrison RP. Vaccination against Chlamydia genital infection utilizing the murine *C. muridarum* model. *Infect Immun*. 2011; 79(3):986–996. doi: IAI.00881-10 [pii] doi: [10.1128/IAI.00881-10](https://doi.org/10.1128/IAI.00881-10) PMID: [21078844](https://pubmed.ncbi.nlm.nih.gov/21078844/); pmcid: PMC3067520.
13. Hawkins RA, Rank RG, Kelly KA. A Chlamydia trachomatis-specific Th2 clone does not provide protection against a genital infection and displays reduced trafficking to the infected genital mucosa. *Infect Immun*. 2002; 70(9):5132–5139. PMID: [12183563](https://pubmed.ncbi.nlm.nih.gov/12183563/); pmcid: PMC128225.
14. Igietseme JU, Rank RG. Susceptibility to reinfection after a primary chlamydial genital infection is associated with a decrease of antigen-specific T cells in the genital tract. *Infect Immun*. 1991; 59(4):1346–1351. PMID: [1900810](https://pubmed.ncbi.nlm.nih.gov/1900810/); pmcid: PMC257849.
15. Karunakaran KP, Yu H, Foster LJ, Brunham RC. Development of a Chlamydia trachomatis T cell Vaccine. *Hum Vaccin*. 2010; 6(8):676–680. doi: 12299 [pii]. PMID: [20523121](https://pubmed.ncbi.nlm.nih.gov/20523121/); pmcid: PMC3056063.
16. Li W, Murthy AK, Guentzel MN, Seshu J, Forsthuber TG, Zhong G, et al. Antigen-specific CD4+ T cells produce sufficient IFN-gamma to mediate robust protective immunity against genital Chlamydia muridarum infection. *J Immunol*. 2008; 180(5):3375–3382. doi: 180/5/3375 [pii]. PMID: [18292563](https://pubmed.ncbi.nlm.nih.gov/18292563/).
17. Li LX, McSorley SJ. B cells enhance antigen-specific CD4 T cell priming and prevent bacteria dissemination following Chlamydia muridarum genital tract infection. *PLoS Pathog*. 2013; 9(10):e1003707. doi: [10.1371/journal.ppat.1003707](https://doi.org/10.1371/journal.ppat.1003707) PPATHOGENS-D-13-01491 [pii]. PMID: [24204262](https://pubmed.ncbi.nlm.nih.gov/24204262/); pmcid: PMC3814678.
18. Morrison RP, Feilzer K, Tumas DB. Gene knockout mice establish a primary protective role for major histocompatibility complex class II-restricted responses in Chlamydia trachomatis genital tract infection. *Infect Immun*. 1995; 63(12):4661–4668. PMID: [7591120](https://pubmed.ncbi.nlm.nih.gov/7591120/); pmcid: PMC173669.
19. Morrison SG, Su H, Caldwell HD, Morrison RP. Immunity to murine Chlamydia trachomatis genital tract reinfection involves B cells and CD4(+) T cells but not CD8(+) T cells. *Infect Immun*. 2000; 68(12):6979–6987. PMID: [11083822](https://pubmed.ncbi.nlm.nih.gov/11083822/); pmcid: PMC97807.
20. Morrison SG, Morrison RP. A predominant role for antibody in acquired immunity to chlamydial genital tract reinfection. *J Immunol*. 2005; 175(11):7536–7542. doi: 175/11/7536 [pii]. PMID: [16301662](https://pubmed.ncbi.nlm.nih.gov/16301662/); pmcid: PMC3514507.
21. Perry LL, Feilzer K, Caldwell HD. Immunity to Chlamydia trachomatis is mediated by T helper 1 cells through IFN-gamma-dependent and-independent pathways. *J Immunol*. 1997; 158(7):3344–3352. PMID: [9120292](https://pubmed.ncbi.nlm.nih.gov/9120292/).
22. Su H, Caldwell HD. CD4+ T cells play a significant role in adoptive immunity to Chlamydia trachomatis infection of the mouse genital tract. *Infect Immun*. 1995; 63(9):3302–3308. PMID: [7642259](https://pubmed.ncbi.nlm.nih.gov/7642259/); pmcid: PMC173455.
23. Yang X, Brunham RC. Gene knockout B cell-deficient mice demonstrate that B cells play an important role in the initiation of T cell responses to Chlamydia trachomatis (mouse pneumonitis) lung infection. *J Immunol*. 1998; 161(3):1439–1446. PMID: [9686609](https://pubmed.ncbi.nlm.nih.gov/9686609/).
24. Hafner LM, Wilson DP, Timms P. Development status and future prospects for a vaccine against Chlamydia trachomatis infection. *Vaccine*. 2014; 32(14):1563–1571. doi: S0264-410X(13)01111-0 [pii] doi: [10.1016/j.vaccine.2013.08.020](https://doi.org/10.1016/j.vaccine.2013.08.020) PMID: [23973245](https://pubmed.ncbi.nlm.nih.gov/23973245/).
25. Karunakaran KP, Yu H, Jiang X, Chan Q, Moon KM, Foster LJ, et al. Outer membrane proteins preferentially load MHC class II peptides: Implications for a Chlamydia trachomatis T cell vaccine. *Vaccine*. 2015; 33(18):2159–2166. doi: S0264-410X(15)00244-3 [pii] doi: [10.1016/j.vaccine.2015.02.055](https://doi.org/10.1016/j.vaccine.2015.02.055) PMID: [25738816](https://pubmed.ncbi.nlm.nih.gov/25738816/); pmcid: PMC4390527.
26. Johnson RM, Yu H, Kerr MS, Slaven JE, Karunakaran KP, Brunham RC. PmpG303-311, a protective vaccine epitope that elicits persistent cellular immune responses in Chlamydia muridarum-immune mice. *Infect Immun*. 2012; 80(6):2204–2211. doi: IAI.06339-11 [pii] doi: [10.1128/IAI.06339-11](https://doi.org/10.1128/IAI.06339-11) PMID: [22431650](https://pubmed.ncbi.nlm.nih.gov/22431650/); pmcid: PMC3370596.
27. Tan C, Hsia RC, Shou H, Haggerty CL, Ness RB, Gaydos CA, et al. Chlamydia trachomatis-infected patients display variable antibody profiles against the nine-member polymorphic membrane protein family. *Infect Immun*. 2009; 77(8):3218–3226. doi: IAI.01566-08 [pii] doi: [10.1128/IAI.01566-08](https://doi.org/10.1128/IAI.01566-08) PMID: [19487469](https://pubmed.ncbi.nlm.nih.gov/19487469/); pmcid: PMC2715660.

28. Yu H, Jiang X, Shen C, Karunakaran KP, Brunham RC. Novel *Chlamydia muridarum* T cell antigens induce protective immunity against lung and genital tract infection in murine models. *J Immunol*. 2009; 182(3):1602–1608. doi: 182/3/1602 [pii]. PMID: [19155509](#); pmcid: PMC2637473.
29. Yu H, Karunakaran KP, Jiang X, Shen C, Andersen P, Brunham RC. *Chlamydia muridarum* T cell antigens and adjuvants that induce protective immunity in mice. *Infect Immun*. 2012; 80(4):1510–1518. doi: IAI.06338-11 [pii] doi: [10.1128/IAI.06338-11](#) PMID: [22290151](#); pmcid: PMC3318408.
30. Karunakaran KP, Rey-Ladino J, Stoynov N, Berg K, Shen C, Jiang X, et al. Immunoproteomic discovery of novel T cell antigens from the obligate intracellular pathogen *Chlamydia*. *J Immunol*. 2008; 180(4): 2459–2465. doi: 180/4/2459 [pii]. PMID: [18250455](#).
31. Henderson IR, Lam AC. Polymorphic proteins of *Chlamydia* spp.—autotransporters beyond the Proteobacteria. *Trends Microbiol*. 2001; 9(12):573–578. doi: S0966-842X(01)02234-X [pii]. PMID: [11728862](#).
32. Henderson IR, Navarro-Garcia F, Desvaux M, Fernandez RC, Ala'Aldeen D. Type V protein secretion pathway: the autotransporter story. *Microbiol Mol Biol Rev*. 2004; 68(4):692–744. doi: 68/4/692 [pii] doi: [10.1128/MMBR.68.4.692-744.2004](#) PMID: [15590781](#); pmcid: PMC539010.
33. Grimwood J, Stephens RS. Computational analysis of the polymorphic membrane protein superfamily of *Chlamydia trachomatis* and *Chlamydia pneumoniae*. *Microb Comp Genomics*. 1999; 4(3):187–201. PMID: [10587946](#).
34. Becker E, Hegemann JH. All subtypes of the Pmp adhesin family are implicated in chlamydial virulence and show species-specific function. *Microbiologyopen*. 2014; 3(4):544–556. doi: [10.1002/mbo3.186](#) PMID: [24985494](#); pmcid: PMC4287181.
35. Kiselev AO, Skinner MC, Lampe MF. Analysis of *pmpD* expression and PmpD post-translational processing during the life cycle of *Chlamydia trachomatis* serovars A, D, and L2. *PLoS One*. 2009; 4(4): e5191. doi: [10.1371/journal.pone.0005191](#) PMID: [19367336](#); pmcid: PMC2666266.
36. Saka HA, Thompson JW, Chen YS, Kumar Y, Dubois LG, Moseley MA, et al. Quantitative proteomics reveals metabolic and pathogenic properties of *Chlamydia trachomatis* developmental forms. *Mol Microbiol*. 2011; 82(5):1185–1203. doi: [10.1111/j.1365-2958.2011.07877.x](#) PMID: [22014092](#); pmcid: PMC3225693.
37. Swanson KA, Taylor LD, Frank SD, Sturdevant GL, Fischer ER, Carlson JH, et al. *Chlamydia trachomatis* polymorphic membrane protein D is an oligomeric autotransporter with a higher-order structure. *Infect Immun*. 2009; 77(1):508–516. doi: IAI.01173-08 [pii] doi: [10.1128/IAI.01173-08](#) PMID: [19001072](#); pmcid: PMC2612253.
38. Crane DD, Carlson JH, Fischer ER, Bavoil P, Hsia RC, Tan C, et al. *Chlamydia trachomatis* polymorphic membrane protein D is a species-common pan-neutralizing antigen. *Proc Natl Acad Sci U S A*. 2006; 103(6):1894–1899. doi: 0508983103 [pii] doi: [10.1073/pnas.0508983103](#) PMID: [16446444](#); pmcid: PMC1413641.
39. Borges V, Nunes A, Ferreira R, Borrego MJ, Gomes JP. Directional Evolution of *Chlamydia trachomatis* towards Niche-Specific Adaptation. *J Bacteriol*. 2012; 194(22):6143–6153. doi: JB.01291-12 [pii] doi: [10.1128/JB.01291-12](#) PMID: [22961851](#); pmcid: PMC3486361.
40. Carlson JH, Porcella SF, McClarty G, Caldwell HD. Comparative genomic analysis of *Chlamydia trachomatis* oculotropic and genitotropic strains. *Infect Immun*. 2005; 73(10):6407–6418. doi: 73/10/6407 [pii] doi: [10.1128/IAI.73.10.6407-6418.2005](#) PMID: [16177312](#); pmcid: PMC1230933.
41. Tan C, Hsia RC, Shou H, Carrasco JA, Rank RG, Bavoil PM. Variable expression of surface-exposed polymorphic membrane proteins in in vitro-grown *Chlamydia trachomatis*. *Cell Microbiol*. 2010; 12(2): 174–187. doi: CMI1389 [pii] doi: [10.1111/j.1462-5822.2009.01389.x](#) PMID: [19811502](#).
42. Borges V, Pinheiro M, Vieira L, Sampaio DA, Nunes A, Borrego MJ, et al. Complete Genome Sequence of *Chlamydia trachomatis* Ocular Serovar C Strain TW-3. *Genome Announc*. 2014; 2(1). doi: 2/1/e01204-13 [pii] doi: [10.1128/genomeA.01204-13](#) PMID: [24459269](#); pmcid: PMC3900901.
43. Harris SR, Clarke IN, Seth-Smith HM, Solomon AW, Cutcliffe LT, Marsh P, et al. Whole-genome analysis of diverse *Chlamydia trachomatis* strains identifies phylogenetic relationships masked by current clinical typing. *Nat Genet*. 2012; 44(4):413–419. S411. doi: ng.2214 [pii] doi: [10.1038/ng.2214](#) PMID: [22406642](#); pmcid: PMC3378690.
44. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One*. 2010; 5(6):e11147. doi: [10.1371/journal.pone.0011147](#) PMID: [20593022](#); pmcid: PMC2892488.
45. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 2009; 25(11):1451–1452. doi: btp187 [pii] doi: [10.1093/bioinformatics/btp187](#) PMID: [19346325](#).
46. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987; 4(4):406–425. PMID: [3447015](#).

47. Nunes A, Nogueira PJ, Borrego MJ, Gomes JP. Chlamydia trachomatis diversity viewed as a tissue-specific coevolutionary arms race. *Genome Biol.* 2008; 9(10):R153. doi: [10.1186/gb-2008-9-10-r153](https://doi.org/10.1186/gb-2008-9-10-r153) [pii] doi: [10.1186/gb-2008-9-10-r153](https://doi.org/10.1186/gb-2008-9-10-r153) PMID: [18947394](https://pubmed.ncbi.nlm.nih.gov/18947394/); pmcid: PMC2760880.
48. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol.* 1980; 16(2):111–120. PMID: [7463489](https://pubmed.ncbi.nlm.nih.gov/7463489/).
49. Nei M, Kumar S. *Molecular Evolution and Phylogenetics.* New York: Oxford University Press; 2000.
50. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol.* 1999; 73(1):152–160. PMID: [9847317](https://pubmed.ncbi.nlm.nih.gov/9847317/); pmcid: PMC103818.
51. Pride DT. SWAAP Version 1.0.0—Sliding Windows Alignment Analysis Program: A tool for analyzing patterns of substitutions and similarity in multiple alignments. Distributed by the author. 2000.
52. Wehrl W, Brinkmann V, Jungblut PR, Meyer TF, Szczepek AJ. From the inside out—processing of the Chlamydial autotransporter PmpD and its role in bacterial adhesion and activation of human host cells. *Mol Microbiol.* 2004; 51(2):319–334. doi: [10.1046/j.1365-2958.2003.03838.x](https://doi.org/10.1046/j.1365-2958.2003.03838.x) [MMI3838](https://pubmed.ncbi.nlm.nih.gov/14756775/) [pii]. PMID: [14756775](https://pubmed.ncbi.nlm.nih.gov/14756775/).
53. Emini EA, Hughes JV, Perlow DS, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol.* 1985; 55(3):836–839. PMID: [2991600](https://pubmed.ncbi.nlm.nih.gov/2991600/); pmcid: PMC255070.
54. Jameson BA, Wolf H. The antigenic index: a novel algorithm for predicting antigenic determinants. *Comput Appl Biosci.* 1988; 4(1):181–186. PMID: [2454713](https://pubmed.ncbi.nlm.nih.gov/2454713/).
55. Henderson IR, Owen P, Nataro JP. Molecular switches—the ON and OFF of bacterial phase variation. *Mol Microbiol.* 1999; 33(5):919–932. doi: [10.1046/j.1365-2958.1999.01555.x](https://doi.org/10.1046/j.1365-2958.1999.01555.x) [pii]. PMID: [10476027](https://pubmed.ncbi.nlm.nih.gov/10476027/).
56. Moxon R, Bayliss C, Hood D. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu Rev Genet.* 2006; 40:307–333. doi: [10.1146/annurev.genet.40.110405.090442](https://doi.org/10.1146/annurev.genet.40.110405.090442) PMID: [17094739](https://pubmed.ncbi.nlm.nih.gov/17094739/).
57. van der Woude MW, Baumler AJ. Phase and antigenic variation in bacteria. *Clin Microbiol Rev.* 2004; 17(3):581–611, table of contents. doi: [10.1128/CMR.17.3.581-611.2004](https://doi.org/10.1128/CMR.17.3.581-611.2004) [17/3/581](https://pubmed.ncbi.nlm.nih.gov/15258095/) [pii]. PMID: [15258095](https://pubmed.ncbi.nlm.nih.gov/15258095/); pmcid: PMC452554.
58. Chelladurai BS, Li H, Nicholson AW. A conserved sequence element in ribonuclease III processing signals is not required for accurate in vitro enzymatic cleavage. *Nucleic Acids Res.* 1991; 19(8):1759–1766. PMID: [1709490](https://pubmed.ncbi.nlm.nih.gov/1709490/); pmcid: PMC328101.
59. Ehretsmann CP, Carpousis AJ, Krisch HM. Specificity of Escherichia coli endoribonuclease RNase E: in vivo and in vitro analysis of mutants in a bacteriophage T4 mRNA processing site. *Genes Dev.* 1992; 6(1):149–159. PMID: [1730408](https://pubmed.ncbi.nlm.nih.gov/1730408/).
60. Robertson HD, Dickson E, Dunn JJ. A nucleotide sequence from a ribonuclease III processing site in bacteriophage T7 RNA. *Proc Natl Acad Sci U S A.* 1977; 74(3):822–826. PMID: [265576](https://pubmed.ncbi.nlm.nih.gov/265576/); pmcid: PMC430490.
61. Deutscher MP. Degradation of RNA in bacteria: comparison of mRNA and stable RNA. *Nucleic Acids Res.* 2006; 34(2):659–666. doi: [10.1093/nar/gkj472](https://doi.org/10.1093/nar/gkj472) [pii] doi: [10.1093/nar/gkj472](https://doi.org/10.1093/nar/gkj472) PMID: [16452296](https://pubmed.ncbi.nlm.nih.gov/16452296/); pmcid: PMC1360286.
62. Shine J, Dalgarno L. The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci U S A.* 1974; 71(4):1342–1346. PMID: [4598299](https://pubmed.ncbi.nlm.nih.gov/4598299/); pmcid: PMC388224.
63. Allen JE, Stephens RS. Identification by sequence analysis of two-site posttranslational processing of the cysteine-rich outer membrane protein 2 of Chlamydia trachomatis serovar L2. *J Bacteriol.* 1989; 171(1):285–291. PMID: [2914847](https://pubmed.ncbi.nlm.nih.gov/2914847/); pmcid: PMC209584.
64. Danilition SL, Maclean IW, Peeling R, Winston S, Brunham RC. The 75-kilodalton protein of Chlamydia trachomatis: a member of the heat shock protein 70 family? *Infect Immun.* 1990; 58(1):189–196. PMID: [2294048](https://pubmed.ncbi.nlm.nih.gov/2294048/); pmcid: PMC258428.
65. de la Maza LM, Fielder TJ, Carlson EJ, Markoff BA, Peterson EM. Sequence diversity of the 60-kilodalton protein and of a putative 15-kilodalton protein between the trachoma and lymphogranuloma venereum biovars of Chlamydia trachomatis. *Infect Immun.* 1991; 59(3):1196–1201. PMID: [1997423](https://pubmed.ncbi.nlm.nih.gov/1997423/); pmcid: PMC258390.
66. Stephens RS, Wagar EA, Edman U. Developmental regulation of tandem promoters for the major outer membrane protein gene of Chlamydia trachomatis. *J Bacteriol.* 1988; 170(2):744–750. PMID: [2448291](https://pubmed.ncbi.nlm.nih.gov/2448291/); pmcid: PMC210717.
67. Nunes A, Gomes JP. Evolution, phylogeny, and molecular epidemiology of Chlamydia. *Infect Genet Evol.* 2014; 23:49–64. doi: [10.1016/j.meegid.2014.01.029](https://doi.org/10.1016/j.meegid.2014.01.029) [pii] doi: [10.1016/j.meegid.2014.01.029](https://doi.org/10.1016/j.meegid.2014.01.029) PMID: [24509351](https://pubmed.ncbi.nlm.nih.gov/24509351/).

68. Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, Aravind L, et al. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science*. 1998; 282(5389):754–759. PMID: [9784136](#).
69. Ferreira R, Antelo M, Nunes A, Borges V, Damiao V, Borrego MJ, et al. In silico scrutiny of genes revealing phylogenetic congruence with clinical prevalence or tropism properties of *Chlamydia trachomatis* strains. *G3 (Bethesda)*. 2014; 5(1):9–19. doi: [g3.114.015354](#) [pii] doi: [10.1534/g3.114.015354](#) PMID: [25378473](#); pmcid: PMC4291473.
70. Finco O, Frigimelica E, Buricchi F, Petracca R, Galli G, Faenzi E, et al. Approach to discover T- and B-cell antigens of intracellular pathogens applied to the design of *Chlamydia trachomatis* vaccines. *Proc Natl Acad Sci U S A*. 2011; 108(24):9969–9974. doi: [1101756108](#) [pii] doi: [10.1073/pnas.1101756108](#) PMID: [21628568](#); pmcid: PMC3116399.
71. Jorgensen I, Valdivia RH. Pmp-like proteins Pls1 and Pls2 are secreted into the lumen of the *Chlamydia trachomatis* inclusion. *Infect Immun*. 2008; 76(9):3940–3950. doi: [10.1128/IAI.00632-08](#) [pii] doi: [10.1128/IAI.00632-08](#) PMID: [18591235](#); pmcid: PMC2519427.
72. Sisko JL, Spaeth K, Kumar Y, Valdivia RH. Multifunctional analysis of *Chlamydia*-specific genes in a yeast expression system. *Mol Microbiol*. 2006; 60(1):51–66. doi: [MMI5074](#) [pii] doi: [10.1111/j.1365-2958.2006.05074.x](#) PMID: [16556220](#).
73. Gomes JP, Nunes A, Bruno WJ, Borrego MJ, Florindo C, Dean D. Polymorphisms in the nine polymorphic membrane proteins of *Chlamydia trachomatis* across all serovars: evidence for serovar Da recombination and correlation with tissue tropism. *J Bacteriol*. 2006; 188(1):275–286. doi: [188/1/275](#) [pii] doi: [10.1128/JB.188.1.275-286.2006](#) PMID: [16352844](#); pmcid: PMC1317584.
74. Borges V, Gomes JP. Deep comparative genomics among *Chlamydia trachomatis* lymphogranuloma venereum isolates highlights genes potentially involved in pathoadaptation. *Infect Genet Evol*. 2015; 32:74–88. doi: [S1567-1348\(15\)00069-6](#) [pii] doi: [10.1016/j.meegid.2015.02.026](#) PMID: [25745888](#).
75. Gomes JP, Bruno WJ, Nunes A, Santos N, Florindo C, Borrego MJ, et al. Evolution of *Chlamydia trachomatis* diversity occurs by widespread interstrain recombination involving hotspots. *Genome Res*. 2007; 17(1):50–60. doi: [gr.5674706](#) [pii] doi: [10.1101/gr.5674706](#) PMID: [17090662](#); pmcid: PMC1716266.
76. Rockey DD, Lenart J, Stephens RS. Genome sequencing and our understanding of chlamydiae. *Infect Immun*. 2000; 68(10):5473–5479. PMID: [10992442](#); pmcid: PMC101494.
77. Feher VA, Randall A, Baldi P, Bush RM, de la Maza LM, Amaro RE. A 3-dimensional trimeric beta-barrel model for *Chlamydia* MOMP contains conserved and novel elements of Gram-negative bacterial porins. *PLoS One*. 2013; 8(7):e68934. doi: [10.1371/journal.pone.0068934](#) [pii] doi: [10.1371/journal.pone.0068934](#) PMID: [23935908](#); pmcid: PMC3723809.
78. Nunes A, Nogueira PJ, Borrego MJ, Gomes JP. Adaptive evolution of the *Chlamydia trachomatis* dominant antigen reveals distinct evolutionary scenarios for B- and T-cell epitopes: worldwide survey. *PLoS One*. 2010; 5(10). doi: [10.1371/journal.pone.0013171](#) PMID: [20957150](#); pmcid: PMC2950151.
79. Rodriguez-Maranon MJ, Bush RM, Peterson EM, Schirmer T, de la Maza LM. Prediction of the membrane-spanning beta-strands of the major outer membrane protein of *Chlamydia*. *Protein Sci*. 2002; 11(7):1854–1861. doi: [10.1110/ps.3650102](#) PMID: [12070338](#); pmcid: PMC2373662.
80. Molleken K, Schmidt E, Hegemann JH. Members of the Pmp protein family of *Chlamydia pneumoniae* mediate adhesion to human cells via short repetitive peptide motifs. *Mol Microbiol*. 2010; 78(4):1004–1017. doi: [10.1111/j.1365-2958.2010.07386.x](#) PMID: [21062373](#); pmcid: PMC2997323.
81. Jeffrey BM, Suchland RJ, Eriksen SG, Sandoz KM, Rockey DD. Genomic and phenotypic characterization of in vitro-generated *Chlamydia trachomatis* recombinants. *BMC Microbiol*. 2013; 13:142. doi: [1471-2180-13-142](#) [pii] doi: [10.1186/1471-2180-13-142](#) PMID: [23786423](#); pmcid: PMC3703283.
82. Grimwood J, Olinger L, Stephens RS. Expression of *Chlamydia pneumoniae* polymorphic membrane protein family genes. *Infect Immun*. 2001; 69(4):2383–2389. doi: [10.1128/IAI.69.4.2383-2389.2001](#) PMID: [11254597](#); pmcid: PMC98169.
83. Read TD, Brunham RC, Shen C, Gill SR, Heidelberg JF, White O, et al. Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Res*. 2000; 28(6):1397–1406. doi: [gkd254](#) [pii]. PMID: [10684935](#); pmcid: PMC111046.
84. Shirai M, Hirakawa H, Ouchi K, Tabuchi M, Kishi F, Kimoto M, et al. Comparison of outer membrane protein genes omp and pmp in the whole genome sequences of *Chlamydia pneumoniae* isolates from Japan and the United States. *J Infect Dis*. 2000; 181 Suppl 3:S524–527. doi: [JID991374](#) [pii] doi: [10.1086/315616](#) PMID: [10839753](#).
85. Tan M. Regulation of gene expression. In: Bavoil P, Wyrick PB, editors. *Chlamydia: genomics and pathogenesis* Horizon Bioscience; 2006.

86. Albrecht M, Sharma CM, Reinhardt R, Vogel J, Rudel T. Deep sequencing-based discovery of the *Chlamydia trachomatis* transcriptome. *Nucleic Acids Res.* 2010; 38(3):868–877. doi: [gkp1032](https://doi.org/10.1093/nar/gkp1032) [pii] doi: [10.1093/nar/gkp1032](https://doi.org/10.1093/nar/gkp1032) PMID: [19923228](https://pubmed.ncbi.nlm.nih.gov/19923228/); pmcid: PMC2817459.
87. Nunes A, Gomes JP, Mead S, Florindo C, Correia H, Borrego MJ, et al. Comparative expression profiling of the *Chlamydia trachomatis* pmp gene family for clinical and reference strains. *PLoS One.* 2007; 2(9):e878. doi: [10.1371/journal.pone.0000878](https://doi.org/10.1371/journal.pone.0000878) PMID: [17849007](https://pubmed.ncbi.nlm.nih.gov/17849007/); pmcid: PMC1963315.
88. Yu H, Karunakaran KP, Jiang X, Brunham RC. Evaluation of a multisubunit recombinant polymorphic membrane protein and major outer membrane protein T cell vaccine against *Chlamydia muridarum* genital infection in three strains of mice. *Vaccine.* 2014; 32(36):4672–4680. doi: [S0264-410X\(14\)00791-9](https://doi.org/10.1016/j.vaccine.2014.06.002) [pii] doi: [10.1016/j.vaccine.2014.06.002](https://doi.org/10.1016/j.vaccine.2014.06.002) PMID: [24992718](https://pubmed.ncbi.nlm.nih.gov/24992718/); pmcid: PMC4148050.