

## Comparative Genomic Analysis of *Chlamydia trachomatis* Oculotropic and Genitotropic Strains†

John H. Carlson,<sup>1</sup> Stephen F. Porcella,<sup>2</sup> Grant McClarty,<sup>3</sup> and Harlan D. Caldwell<sup>1\*</sup>

Laboratory of Intracellular Parasites<sup>1</sup> and Laboratory of Human Bacterial Pathogenesis,<sup>2</sup> Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health, 903 South 4th Street, Hamilton, Montana, 59840, and Department of Medical Microbiology, University of Manitoba, 730 William Avenue, Winnipeg, Manitoba, Canada, R3E 0W3<sup>3</sup>

Received 31 March 2005/Returned for modification 16 May 2005/Accepted 2 June 2005

***Chlamydia trachomatis* infection is an important cause of preventable blindness and sexually transmitted disease (STD) in humans. *C. trachomatis* exists as multiple serovariants that exhibit distinct organotropism for the eye or urogenital tract. We previously reported tissue-tropic correlations with the presence or absence of a functional tryptophan synthase and a putative GTPase-inactivating domain of the chlamydial toxin gene. This suggested that these genes may be the primary factors responsible for chlamydial disease organotropism. To test this hypothesis, the genome of an oculotropic trachoma isolate (A/HAR-13) was sequenced and compared to the genome of a genitotropic (D/UW-3) isolate. Remarkably, the genomes share 99.6% identity, supporting the conclusion that a functional tryptophan synthase enzyme and toxin might be the principal virulence factors underlying disease organotropism. Tarp (translocated actin-recruiting phosphoprotein) was identified to have variable numbers of repeat units within the N and C portions of the protein. A correlation exists between lymphogranuloma venereum serovars and the number of N-terminal repeats. Single-nucleotide polymorphism (SNP) analysis between the two genomes highlighted the minimal genetic variation. A disproportionate number of SNPs were observed within some members of the polymorphic membrane protein (*pmp*) autotransporter gene family that corresponded to predicted T-cell epitopes that bind HLA class I and II alleles. These results implicate Pmps as novel immune targets, which could advance future chlamydial vaccine strategies. Lastly, a novel target for PCR diagnostics was discovered that can discriminate between ocular and genital strains. This discovery will enhance epidemiological investigations in nations where both trachoma and chlamydial STD are endemic.**

*Chlamydia trachomatis* isolates exist as 15 serovariants that are separated according to pathobiotypes: trachoma or lymphogranuloma venereum (LGV). Trachoma biovariants consist of serovars A, B, Ba, C, D, E, F, G, H, I, J, and K. LGV biovariants consist of serovars L1, L2, and L3. Trachoma biovariants are noninvasive, epitheliotropic strains that cause blinding trachoma (A to C) or sexually transmitted diseases (STDs) (D to K) (36). LGV serovars cause sexually transmitted disease with disseminating infection of draining regional lymph nodes (37). Although the ocular and genital serovars are capable of infecting epithelial cells of both the conjunctivae and genital tract (38), they exhibit distinct tropisms in terms of organ-specific disease that we have previously termed organotropism.

The genome of the *C. trachomatis* STD strain D/UW-3 has been publicly available since 1998 (39). DNA microarray studies, in which test serovar DNAs were hybridized against target D/UW-3 sequences, demonstrated that the *C. trachomatis* genomes are strikingly similar to each other and are estimated to share greater than 99% identity (3, 8). Genetic differences observed by DNA microarray analysis centered in the 50-kb plasticity zone, *ompA* (major outer membrane protein

[MOMP]), and members of the polymorphic membrane protein (*pmp*) gene family. Genetic variation in *ompA* is primarily located in four variable domains (47). The variable domains are surface accessible and immunodominant and elicit antibodies that divide the strains into the 15 serovariants (5, 40). To date, there is no correlation between MOMP serovariation and pathobiotypes (4, 36).

Previous attempts to determine the underlying genetic basis for ocular or genital tissue tropism in *C. trachomatis* serovars have focused on region-specific comparative genomics (8, 12, 41). STD serovars retain a functional *trpBA* operon encoding tryptophan synthase, whereas ocular serovars have accumulated mutations in the *trpBA* genes that inactivate the enzyme (12). In addition, genital, but not ocular, serovars possess an intact open reading frame (ORF) (CT166) that contains a putative GTPase-inactivating domain of unknown function (8). Finally, sequence variation within *pmpH* distinguishes between ocular, genital, and LGV pathotypes (41).

While a region-specific analysis of all *C. trachomatis* serovars has provided new data, it is by no means comprehensive. To definitively elucidate the genetic basis that defines *C. trachomatis* ocular and genital pathodiversity, we sequenced the genome of a plaque-cloned ocular trachoma serovar (A/HAR-13) isolate. Direct comparative genomic analysis was performed against the closely related, but pathobiologically distinct, urogenital strain D/UW-3. Our results define for the first time the genetic differences between two serovars that cause widely different diseases in the human host.

\* Corresponding author. Mailing address: Laboratory of Intracellular Parasites, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health, 903 South 4th Street, Hamilton, Montana, 59840. Phone: (406) 363-9333. Fax: (406) 363-9380. E-mail: hcaldwell@niaid.nih.gov.

† Supplemental material for this article may be found at <http://iai.asm.org>.

## MATERIALS AND METHODS

**Chlamydiae.** The *C. trachomatis* strains analyzed in this study were A/HAR-13, B/TW-5/OT, Ba/AP-2, C/TW-3/OT, D/UW-3/Cx, E/Bour, F/IC-Cal-13, G/UW-524/Cx, H/UW-4/Cx, I/UW-4/Ur, J/UW-36/Cx, K/UW-31/Cx, L1/LGV-440, L2/LGV-434, and L3/LGV-404. All chlamydiae were grown and purified as previously described (8).

**Genomic DNA purification for sequence analysis.** *C. trachomatis* strain A/HAR-13 was plaque purified on McCoy (ATCC, Manassas, VA) cells (25), and a single clone (clone 6) was grown in sufficient quantities in HeLa 229 cells (ATCC) for purification of elementary bodies (EBs). Genomic DNA was purified from  $7 \times 10^9$  density gradient-purified EBs in the following manner. One-milliliter volumes (5 ml total) of EB suspensions were microcentrifuged at 15,000 rpm for 10 min at 4°C to pellet the cells. The supernatant was aspirated off, and each pellet was suspended in 546  $\mu$ l of TE (50 mM Tris, 50 mM EDTA, pH 8.0), followed by the addition of 45  $\mu$ l 10% sodium dodecyl sulfate, 6  $\mu$ l 1 M dithiothreitol, and 3  $\mu$ l of proteinase K (~18 mg/ml; recombinant; PCR grade; Roche Applied Science, Indianapolis, IN). The suspensions were incubated at 60°C for 2 hours, with an additional 3  $\mu$ l of proteinase K added at 20-min intervals. The solutions were then extracted three times with a 1 $\times$  volume of phenol-chloroform-isoamyl alcohol (25:24:1; Roche Applied Science), followed by two extractions with a 1 $\times$  volume of chloroform-isoamyl alcohol (24:1). The aqueous layer from each tube was placed in a fresh tube with 1/10 volume of 3 M sodium acetate, pH 5.5 (Ambion, Inc., Austin, TX), and 0.6 volume of isopropanol was added. The solutions were immediately microcentrifuged at 8,000 rpm for 10 min at 4°C. After the supernatant was aspirated, the pellets were washed once with cold 70% ethanol, and the pellets were allowed to air dry at room temperature. The pellets were suspended in a total volume of ~900  $\mu$ l of TE (pH 8.0), and UV readings at 260 nm were taken to determine the DNA concentration. Approximately 365  $\mu$ g of DNA was sent to Integrated Genomics, Inc. (Chicago, IL) for genome sequencing.

**Genome sequencing, annotation, and alignment.** The genome of A/HAR-13 was sequenced by methods used for several other bacteria (11, 19, 21). Directed sequencing was performed to increase the minimum consensus base quality to Q40 (99.99% accuracy of base call) for regions of low sequence quality in the assembled genome. ORFs were identified with proprietary software (Integrated Genomics) and via manually focused efforts and were entered into the ERGO bioinformatics suite for final annotation (32). GC skew was calculated as  $(C - G)/(G + C)$ , with a 20-kb sliding window moving in 500-bp incremental steps. GC% was calculated using 2-kb and 20-kb sliding windows compared to total GC% for the entire chromosome.

Protein and DNA alignments were performed using the DNASTAR LaserGene software package (Madison, WI) according to the manufacturer's recommendations. Briefly, DNA or protein sequences were first aligned using ClustalV (18) at default values. The alignment was then transferred into Canvas 8 (ACD Systems, Miami, FL), where alignment features were added or highlighted.

**SNP identification protocol.** Single-nucleotide polymorphisms (SNPs) were identified between *Chlamydia trachomatis* D/UW-3 and A/HAR-13 by first clustering all ORFs, RNAs, and intergenic regions into those respective group types. The DNA types were further filtered using a similarity cutoff of 80% of bases between the two genomes of the same type. If features were no more than 10% different in overall length, then the features were considered clustered together and were used for calculation of SNPs. Features with greater than 10% difference were manually analyzed. ClustalW was used to generate alignments between all DNA sequences in all clusters for each feature type. SNPs were defined where one or more of the aligned sequences had a change in a nucleotide at a specific location in the alignment. Two or more sequentially positioned SNPs were manually analyzed. SNP results were loaded into ERGO for further detailed analysis (32).

**PCR analyses.** The following primer sequences were designed for PCR amplification of the deletions/insertions found in CTA0934 (JHC452, 5'-ATG TAC AGA CAG TTA GAA TCC AAC-3'; JHC453, 5'-CTA GTC AGC GAG TGG GAT GAC-3'), CTA0054 (JHC458, 5'-CGT TAA TCC AAA TAA TAA CAG C-3'; JHC459, 5'-TTA AGG CTG TGA TGG TGT CGA AG-3'), CTA0498 5' deletion (JHC463, CAG AAA GCT CAG AAA CTA GCT CTA CC; JHC464, CCC TCG TTA CGA GGC CCT GTT GTG), and CTA0498 3' deletion (JHC460, 5'-TGC GTA AAC ACC TAG ACG TCG-3'; JHC462, 5'-TGT TTG TAC GCT ACC ACT TCC TCC-3'). Template crude genomic DNAs were prepared as previously described (8). PCR amplifications were conducted using the Expand High Fidelity PCR System (Roche Applied Science) according to the manufacturer's recommendations. All amplification reaction conditions consisted of 35 cycles of 95°C for 30 s, 55°C for 40 s, and 72°C for 1 min, followed by a final 72°C incubation for 5 min. CTA0498 (CT456) amplification conditions

TABLE 1. Comparative genome statistics for *C. trachomatis* oculotropic strain A/HAR-13 and genitotropic strain D/UW-3

Characteristic	Value	
	A/HAR-13	D/UW-3
Chromosome size (bases)	1,044,459	1,042,519
Plasmid size (bases)	7,510	7,493
Genome G+C content (bases)	434,113	433,357
% G+C content	41.267	41.272

varied in that the 5'-deletion annealing temperature was 70°C while the 3'-deletion annealing temperature was 65°C. Amplified products were viewed by electrophoresis of 10- $\mu$ l volumes of each reaction mixture on a 2% agarose-Tris-borate-EDTA gel, followed by staining with ethidium bromide.

**MHC epitope analyses.** The primary sequence of MOMP and PmpF from both A/HAR-13 and D/UW-3 were analyzed using the SYFPEITHI algorithm (33; <http://www.syfpeithi.de>) to identify putative T-cell epitopes recognized by human major histocompatibility complex (MHC) ligands. Using this analysis, naturally presented epitopes should be among the top-scoring 2% of all predicted peptides. We included only those epitopes that gave a score of 25 or greater (maximum score, 36) in our analyses. By this criterion, the greatest number of epitopes for a given haplotype was in HLA-DRB1\*0401(DR4Dw4), with a total of 28 (2.7% of total predicted 15-mer epitopes). The total number of epitopes identified for each of the remaining haplotypes fell below the recommended 2% cutoff.

**Nucleotide sequence accession numbers.** The genome sequence of A/HAR-13 has been deposited in the GenBank database under accession numbers CP000051 (chromosome) and CP000052 (plasmid).

## RESULTS

**Genome characterization.** A total of 920 ORFs were manually annotated in the *C. trachomatis* A/HAR-13 genome, 8 of which are carried on the autonomous plasmid pCTA. Table 1 and Fig. 1 compare the basic genome characteristics of the oculotropic A/HAR-13 to those of the genitotropic *C. trachomatis* D/UW-3. As can be observed, there is a high degree of similarity between the two genomes. Of particular note, the first (outer) circle of Fig. 1 shows the relative location of each A/HAR-13 ORF. The DNA sequence from each ORF was compared to the D/UW-3 genome sequence to identify those ORFs that are shared between the two genomes versus those ORFs that are unique to the A/HAR-13 genome. Only two A/HAR-13 ORFs were identified as having no homolog present in D/UW-3, specifically, CTA0177 and CTA0178. A complete listing of the A/HAR-13 annotation can be found in Table S1 in the supplemental material.

**Genomic differences between A/HAR-13 and D/UW-3.** Table 2 lists all deletions of  $\geq 10$  bp in size, as well as those that are  $\leq 10$  bp and not a multiple of 3, identified by a chromosomal alignment between the two strains. Remarkably, there is a net difference of only 1,940 bp between the A/HAR-13 and D/UW-3 chromosomes; 1,557 bp of this genomic difference is localized within the plasticity zone, a result consistent with previous microarray analyses (3, 8) and genomic comparison of the toxin gene structures from all 15 human chlamydial serovars (8).

A large in-frame deletion was found in CTA0498/CT456, the gene encoding Tarp (translocated actin-recruiting phosphoprotein) (10), and was studied further (see below). We also found a deletion of 125 bp, resulting in an in-frame fusion of two small previously unannotated ORFs present in the

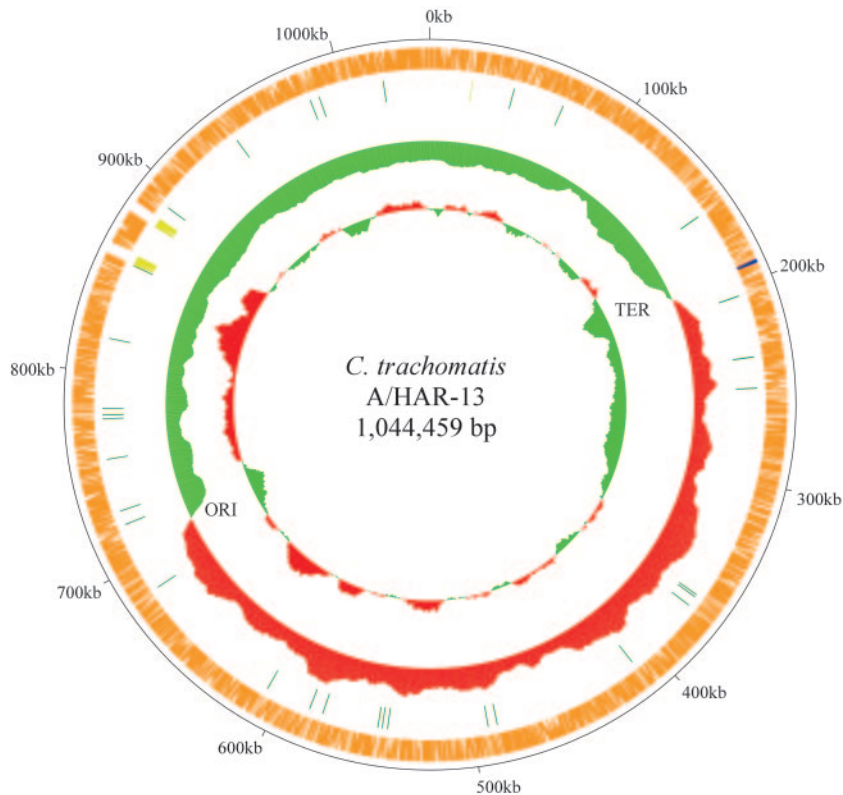


FIG. 1. Circular representation of the A/HAR-13 chromosome. First (outer) circle, alignment of each ORF identified in A/HAR-13 with ORFs in the D/UW-3 chromosomal sequence. Those ORFs that share homology (E value cutoff of  $\geq 1e-10$ ) in D/UW-3 are indicated in orange. Those ORFs lacking a D/UW-3 homolog are indicated in blue. Second circle, location of RNA encoding regions; tRNAs in dark green, rRNAs in yellow. Third circle, light green and red represent GC skew calculated as C/G, 20-kb sliding window, with 500-bp incremental steps. Red and light-green profiles signify regional prevalence of C over G compared to the average CG value for the entire chromosome. Fourth circle, regional GC% (over a 20-kb sliding window) compared to the total 41.27% GC% for the chromosome. Red indicates regions with a GC% above the total average, while light green indicates regions with a GC% below the average. The relative locations of the origin of replication (ORI) and terminus of replication (TER) are also indicated.

D/UW-3 genome. The fusion results in a single ORF (CTA0934) of 330 bp. The function of CTA0934 is unknown, but the deletion and resulting ORF are unique to oculotropic serovars (see below).

Our analysis revealed 8 interrupted ORFs (compared to their D/UW-3 orthologs) in the A/HAR-13 genome (Table 3). We have previously described inactivating mutations present in the *trpRBA* operon (6, 12) and the toxin loci (8) that are unique to oculotropic strains. In addition, strain D/UW-3 possesses two copies of the *tyrP* gene, while the second copy of the *tyrP* gene in A/HAR-13 is disrupted. Similar disrupting mutations in the tandemly repeated *tyrP* loci of *Chlamydia pneumoniae* clonal isolates (14) have been described and functionally linked to differences in tryptophan utilization in vitro. The A/HAR-13 *secD/F* homolog is disrupted by a single-bp deletion, resulting in the separation of the *secD* (CTA0490) and *secF* (CTA0489) genes. The effect this difference has on the functions of SecD and SecF in A/HAR-13 is unknown but may not be biologically significant, as SecD and SecF are also encoded separately in *Escherichia coli* (13). The four remaining gene disruptions identified during our comparative analysis occur in *truB* (tRNA pseudouridine synthase), *arcD* (arginine/ornithine antiporter), and the orthologs of CT105 and CT163,

both encoding hypothetical proteins. The effects of these gene disruptions are unknown.

**CTA0498/CT456 deletional variations among *C. trachomatis* serovars.** Two insertions or deletions unique to CTA0498/CT456 were observed, one of 42 bp and the other 345 bp (Table 2). CTA0498/CT456 encodes the 180-kDa Tarp protein, a type III secreted tyrosine-phosphorylated protein that has actin recruitment properties (10). The CT456 ortholog from the invasive *C. trachomatis* biovar LGV (strain L2/LGV-434) was previously sequenced (10). We aligned the three Tarp proteins, each originating from a serovar representing the three human chlamydial biovariants, and identified further genetic heterogeneity within the proteins (Fig. 2), a result mirrored in the DNA alignment (data not shown). Most notably, the 5' and 3' halves of the CT456 orthologs have undergone extensive deletions or insertions. It appears these 5'-3' polymorphisms have occurred by a recombinational mechanism. The L2/LGV-434 ortholog has six repeat units of ~50 amino acids (aa) located in the 5' insertion/deletion coding region of the gene. In contrast, the A/HAR-13 and D/UW-3 orthologs possess ~3 repeats. Conversely, the 3' half of the A/HAR-13 gene (CTA0498) contains three repeat units, each coding for



TABLE 2. Synopsis of comparative chromosomal deletions between A/HAR-13 and D/UW-3

Deletion location <sup>a</sup>	Deletion size <sup>b</sup> (bp)	Result
CT050	144	In-frame deletion
CTA0122 ( <i>incD</i> )	13	Frameshift; 30 bp added to ORF
CT154/CT155	453	Truncates CTA0163 by 138 bp Truncates CTA0164 by 3 bp
CT156/CT157	158	Intergenic
CTA0175	22	Frameshift; 57 bp added to ORF
CT166/CT167	946	Toxin region <sup>c</sup>
CT326	111	In-frame deletion
CT414 ( <i>pmpC</i> )	42	In-frame deletion
CTA0498 (Tarp)	42	In-frame deletion
CT456 (Tarp)	345	In-frame deletion
CT622	12	In-frame deletion
CT662/CT663	7	Intergenic
CT680/CT681	8	Intergenic
CTA0891/CTA0892	10	Disrupts <i>tyrP_2</i> <sup>d</sup> homolog
CTA0917/CTA0918	7	Intergenic
CTA0934	125	Fuses two small ORFs in D, creating a single, larger ORF in A
CT869 ( <i>pmpE</i> )	12	In-frame deletion
CTA0953	70	Alters ORF annotation between strains

<sup>a</sup> CT numbers refer to D/UW-3-specific deletions and their relative locations, while CTA numbers refer to A/HAR-13 specific deletions and their relative locations. The ORFs flanking intergenic deletions are noted as well.

<sup>b</sup> All deletions larger than 10 bp are listed, as well as any deletion less than 10 bp that is not a multiple of 3.

<sup>c</sup> The annotated ORFs between the two strains are dramatically altered and have been previously reported (8).

<sup>d</sup> *tyrP-2*, second copy of the *tyrP* gene (39).

~120 aa, while D/UW-3 has two repeats, and L2/LGV-434 has a single repeat sequence.

Our initial sequence comparisons (Fig. 2) suggested there might be a correlation between the CT456 sequence polymorphisms and strain organotropism. Therefore, we PCR amplified the regions across either the 5' or 3' deletion region of all 15 serovars. The PCR results of this analysis are shown in Fig. 3. No exact correlation was observed between the number of repeat motifs and a particular pathotropic group. All ocular serovars (A to C) contained 5'-specific regions of similar sizes; however, the genitotropic serovar I shared a product of the same size as the ocular serovars. Moreover, genitotropic serovars tended to vary even further in that serovar E contained fewer repeats than the closely related serovar D/UW-3. Of note, each of the LGV strains appeared to possess a minimum

of 6 repeat elements in the 5' region, while serovars L1 and L3 possessed ~4 additional repeats in comparison to L2. In addition, the 3'-specific repeat region denotes a general pattern of the three genotypes, although again, there is no exact correlation between the number of repeat elements and pathogenic grouping. For example, serovars A, B, and Ba have three repeat elements in the 3' region and C to G and I to K contain two repeat elements, while H and the LGV strains contain only one. Strict definition of the repeat units needs to be confirmed by sequence analysis of the remaining 12 orthologs to confirm our PCR-based analysis.

**SNP analysis.** A total of 3,354 SNPs were identified between the A/HAR-13 and reference D/UW-3 genomes, thus exhibiting 99.6% identity. Nine SNPs localized within RNA encoding regions, 226 in noncoding (intergenic) regions, and 3,119 within coding (ORF) regions. A summary of these results can be found in Table S2 in the supplemental material. Given the limited knowledge of chlamydial gene regulation and promoter sequences, we chose to focus our analysis on SNPs localized within coding regions. Of the 3,119 SNPs identified in ORFs, 1,706 resulted in nonsynonymous amino acid substitutions. Figure 4 shows the distribution of SNPs localized within the 920 ORFs. Remarkably, 73% of the predicted ORFs contain  $\leq 2$  SNPs, further demonstrating the extraordinary degree of identity between the two genomes. There is a distinct clustering of SNPs among ORFs in the A/HAR-13 genome; 46.6% (1,455/3,119) of the SNPs occur in 21 ORFs. These genes, as well as all members of the *pmp* gene family, are ranked according to the total number of SNPs in Table 4.

Genes such as *dppD*, *pyrH*, and *karG* contain a disproportionate number of synonymous SNPs compared to nonsynonymous SNPs. Conversely, *tsf* contains a higher percentage of nonsynonymous SNPs. Since these proteins are essential for cell survival, it is unlikely that the SNPs dramatically alter their respective functions. Five genes encoding hypothetical proteins contain a disproportionately large number of SNPs. Three of these genes are found in a contiguous gene cluster (CTA0053 to CTA0055), an organization suggestive of related gene function. While the functions of these hypothetical proteins are unknown, the high number of SNPs implies that they are under selective pressure and that the proteins could play a role in the pathogenic differentiation separating oculogenital strains.

The *pmp* gene family encodes proteins with homology to *E. coli* autotransporters (15). Stothard et al. (41) first described polymorphisms among the *C. trachomatis pmp* genes by restriction fragment length polymorphism and sequence analysis. While several members of this gene family have been localized to the outer membrane (29, 44), the functions of these proteins remain largely unknown, with the exception of PmpD. In *C. pneumoniae*, the PmpD ortholog was shown to localize to the outer membrane and act as a target of neutralizing antibodies, implicating PmpD as an important virulence factor (46). In this context, it is of interest that some *pmp* genes contain few SNPs (*pmpA*, *pmpD*, *pmpI*, and *pmpG*), while others (*pmpE*, *pmpH*, and *pmpF*) rank among the highest in SNP content between the two genomes.

**MHC predictive epitope mapping of the highly divergent PmpF.** Nonsynonymous SNPs arise from random genetic drift and become predominant in a population only if they result in

TABLE 3. Predicted ORF interruptions in the A/HAR-13 genome compared to D/UW-3

D/UW-3 homolog	Sequence variation	Resulting A/HAR-13 ORFs
CT094 ( <i>truB</i> )	Single nt insertion	CTA0099–CTA0100
CT105	Nonsense mutation	CTA0111–CTA0112
CT163	Single nt deletion	CTA0172–CTA0173
CT167 ( <i>tox</i> )	Single nt deletion	CTA0179–CTA0180
CT170 ( <i>trpB</i> )	Nonsense mutation	CTA0184–CTA0185
CT374 ( <i>arcD</i> )	3 single nt deletions	CTA0406–CTA0408
CT448 ( <i>secD/F</i> )	Single nt deletion	CTA0489–CTA0490
CT818 ( <i>tyrP_2</i> )	10-bp deletion	CTA0891–CTA0892

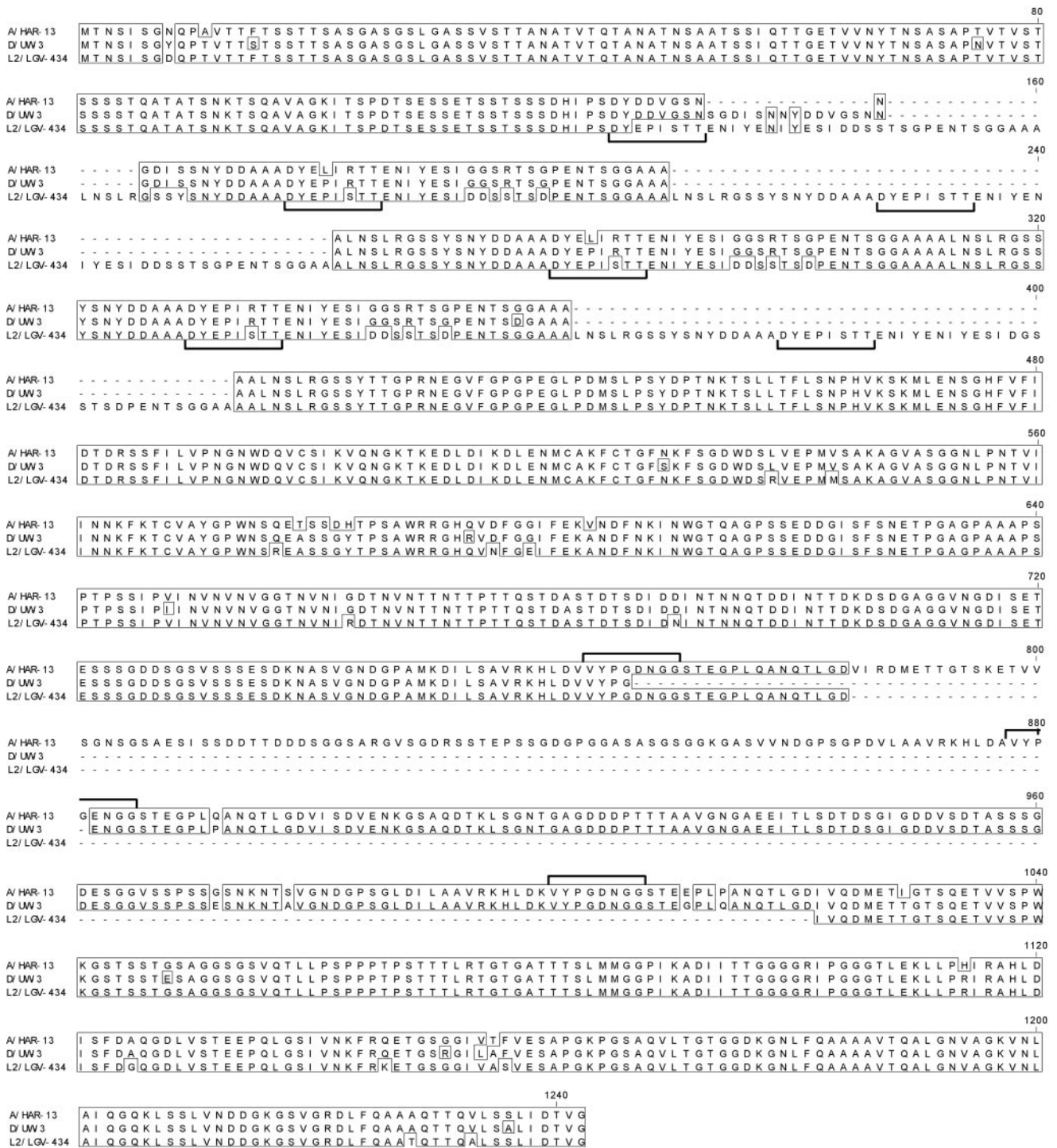


FIG. 2. Alignment of the Tarp orthologs from A/HAR-13, D/UW-3, and L2/LGV-434; ClustalV (18) protein alignment of A/HAR-13 Tarp (this study), D/UW-3 Tarp (39), and L2/LGV-434 Tarp (10). The first 8 aa of the ~50-aa amino-terminal repeat motifs are shown with underlying brackets, and the first 8 aa of the ~120-aa carboxy-terminal repeat motifs are shown with overlying brackets.

a selectable metabolic function, an advantage in pathogen biology, or evasion of host defenses. In the situation of immune evasion, nonsynonymous SNPs cluster within a region(s) of a gene whose protein product is a predominant target of the host immune response. This is exemplified by *ompA*, the gene that

encodes MOMP. MOMP is a well-characterized immunodominant antigenically variable surface antigen recognized by neutralizing antibodies (1, 42, 43, 48), as well as CD4<sup>+</sup> (30, 31) and CD8<sup>+</sup> (22, 23) T cells. *ompA* contains 180 SNPs between A/HAR-13 and D/UW-3, 53 of which are nonsynonymous (Ta-

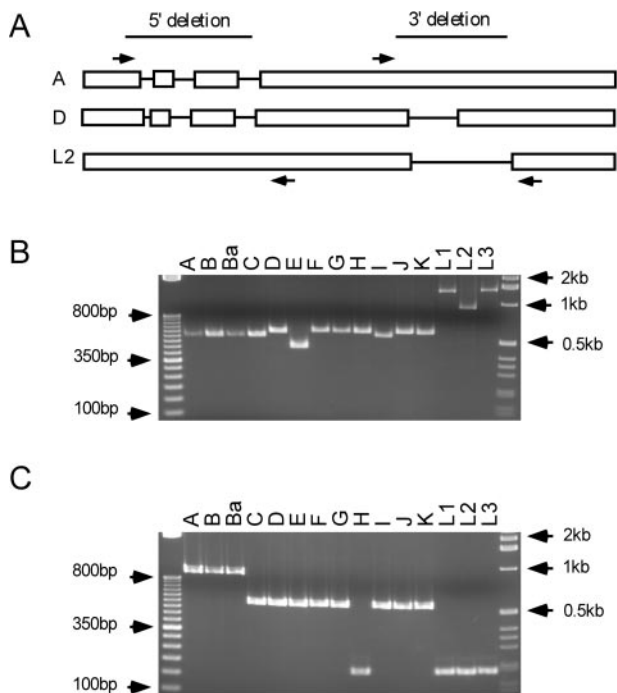


FIG. 3. PCR amplification of the CT456/Tarp 5' and 3' deletions in the 15 *C. trachomatis* reference strains. (A) Schematic representation of the Tarp-encoding genes for serovars A (A/HAR-13), D (D/UW-3), and L2 (L2/LGV-434). The open boxes indicate the coding regions present in the strain of interest (in relation to the strain containing the unaltered sequence), while the lines delineate the deleted region(s). The arrows indicate the relative locations of the PCR primer binding sites. (B) Agarose gel of the PCR amplification results for the 5' deletion region. Predicted product sizes from the known sequences are 565 bp (A/HAR-13), 607 bp (D/UW-3), and 972 bp (L2/LGV-434). A 1-μg 50-bp DNA ladder is loaded in lane 1, and 1-μg 1-kb ladder is loaded in lane 17 (Invitrogen). (C) PCR amplification results for the 3' deletion region. Predicted product sizes for the known sequences are 908 bp (A/HAR-13), 463 bp (D/UW-3), and 197 bp (L2/LGV-434). The size markers and lanes are the same as in panel B.

ble 4). It is therefore noteworthy that *pmpF* exhibits an even greater number of nonsynonymous SNPs than *ompA*, implicating the protein as a potential, heretofore-unrecognized target of the host immune response.

Because of the high number of nonsynonymous SNPs in *pmpF*, we subjected the sequence to the computer-based algorithm SYFPEITHI (33) to ascertain if a correlation exists between nonsynonymous SNP location and predicted MHC class I and II T-cell epitopes. The PmpF primary sequences of the A/HAR-13 and D/UW-3 strains are depicted in Fig. 5. The substitutions primarily cluster in two regions (I and II), and both regions are located within the predicted passenger domain. Region I consists of 87 amino acids, while region II consists of 378 amino acids, representing 15.4% and 71.4% of the total amino acid substitutions (AAS), respectively. We hypothesized that if the localized clustering of AAS is driven by immune selection, then these regions should contain an increased proportion of epitopes recognized by the MHC glycoproteins. We found that 48.5% of the predicted class I and 41.7% of the class II epitopes mapped to region II. These percentages of predicted epitopes are disproportionately high,

since region II constitutes only 36.6% of the protein. Alternatively, region I (representing 8.4% of the protein) exhibits a disproportionate number of class II epitopes (12.5%), but not class I (5.9%) epitopes. Thus, MHC epitope predictions mirror AAS clustering, suggesting that immune selection is driving *pmpF* variation, a finding that indirectly implicates PmpF as a virulence factor and primary target of host cellular immune responses.

**A novel PCR amplification target for the simultaneous detection and discrimination of oculotropic and genitotropic serovars.** As described above, a genomewide scan discovered a difference between the two serovars that was particularly striking, specifically, the 125-bp deletion specific to the A/HAR-13 genome (Table 2). We asked if this CTA0934-specific deletion was unique to oculotropic strains. To test this, primers were designed to screen all 15 *C. trachomatis* reference serovars for the presence or absence of the deletion. Figure 6 shows the results of this analysis and demonstrates that the loss of the 125-bp fragment is unique to oculotropic serovars. In addition, we similarly tested 75 clinical *C. trachomatis* isolates representing both oculotropic and genitotropic serovars and confirmed the oculotropic specificity of this deletion (data not shown). PCR analysis of the CTA0054/CT050-related deletion indicated there was not a similar correlation (data not shown), and the Tarp-specific insertion/deletion results were not as definitive (Fig. 3). Therefore, the CTA0934 marker may prove useful in differentiating between oculotropic and genitotropic serovars.

## DISCUSSION

Microarray (DNA-DNA) analyses suggested that the 15 reference strains of *C. trachomatis* share a degree of identity approaching 99% (3, 8). However, these studies were limited in their interpretation by the fact that the target genome(s) is cross-hybridized to DNA from unsequenced test genomes. Therefore, genes present in the test genome but absent in the target genome will be missed. Moreover, the DNA microarray does not accurately depict genomic SNPs, polymorphisms that can be important to pathogen biology in a background of genomic synteny. Genome sequencing and analysis are therefore the only definitive ways to characterize genetic differences and similarities. To better understand the genetic basis of *C. trachomatis* organotropism, we sequenced the genome of the oculotropic strain A/HAR-13 and compared it to the published sequence of the genitotropic strain D/UW-3. The significance of this work is the finding that the two genomes are 99.6% identical, yet they possess a number of important differences that may be involved in disease organotropism determination.

It is important to note that the sequence of the *C. trachomatis* A/HAR-13 genome was derived from a clonal population originating from a single plaque-purified isolate. Previous chlamydial genomes sequenced used DNA prepared from non-clonal, mixed populations. As such, the sequence derived from the early genome-sequencing projects likely reflected the predominant genotype in the population. In the case of our genome sequence, we can say only that the alleles reflected in our sequence are specific to our clone. For example, it may be that the majority of the A/HAR-13 population contains an intact *trpB* allele, as previously reported (12), but in our isolated clone, it is disrupted. While it is unclear which of the *trpB*



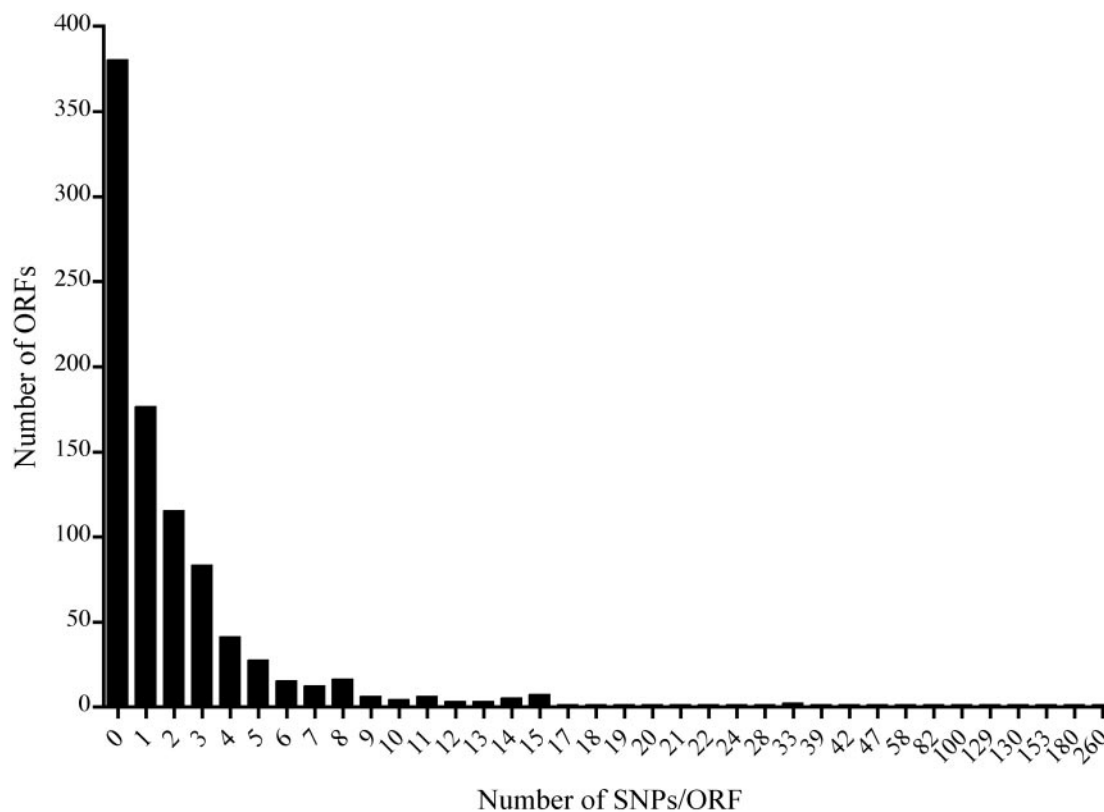


FIG. 4. Distribution of SNPs per ORF in A/HAR-13 versus D/UW-3. The number of SNPs located within a given ORF is indicated on the x axis. The total number of ORFs containing the indicated number of SNPs is indicated on the y axis.

sequences that we have reported is the predominant genotype, the resulting phenotype (tryptophan synthase negative) is the same for both isolates. The same clonal-variation argument could be made for the remaining seven ORFs that were identified as disrupted. Given that the number of genes needing to be targeted is low, further sequence analysis of different plaque-cloned isolates from serovars D and A are warranted to definitively identify the critical mutations.

Outside of SNP-introduced polymorphisms and the previously noted plasticity zone differences (i.e., toxin [8] and tryptophan synthase [12] genes), the number of target genes determining oculogenital infection and disease specificity remains limited. While it is likely that the pathogenic determining factors differentiating between these two strains do not lie within the presence or absence of a single gene, our results strongly suggest that the tryptophan (6, 12) and toxin (8) genotypes play prominent roles in the segregation of oculogenital diseases. Indeed, given the limited differences in gene content and sequence variation, it would be relatively straightforward to target these individual genes in an attempt to genetically transform oculotropic to genitotropic strains, and vice versa. Unfortunately, chlamydiae are genetically intractable organisms, a characteristic that prevents these otherwise straightforward approaches from being applied for defining the molecular basis of *C. trachomatis* pathogenesis.

The most obvious difference between ocular and genital serovars was the level of SNPs observed in a restricted number of genes. The SNPs have important implications in both chla-

mydial biology and immunology. Biologically, intergenic SNPs that lie in promoter regions and ribosome binding sites can have important effects on transcription and translation expression levels. However, little is known about chlamydial promoter or ribosome binding site sequences, preventing predictive analyses from being conducted using sequence data alone. Therefore, more can be deduced from the SNP analysis in regard to immunity and immune evasion. For example, the majority of the *pmp* gene family members (five of nine) were associated with a higher SNP frequency, as were five genes encoding hypothetical proteins. The two genes with the greatest numbers of SNPs were *ompA* (MOMP) and *pmpF*. The high frequency of SNPs resulting in amino acid substitutions in MOMP has been reported previously (47) and is in keeping with its being an antigenically variable immunodominant surface protein that is the target of neutralizing antibodies (43, 48) and T-cell immunity (23, 30).

The polymorphic nature of the *pmp* genes argues that these genes may also be subject to significant immune selection. Particularly noteworthy is our finding that *pmpF* exhibits more extensive SNPs than MOMP, providing indirect evidence that the protein is a dominant target of the host immune response. The functions of Pmp proteins have not been extensively studied. However, all members of the family are transcribed (24), and a subset of the gene products are known to be associated with the chlamydial cell surface (29, 44). The paralogous gene family member, PmpD, has been shown to be surface exposed and processed similarly to other prokaryotic autotransporter

TABLE 4. ORFs containing the highest number of SNPs, as well as all members of the Pmp gene family

ORF <sup>a</sup>	Predicted size (aa)	Total no. of SNPs <sup>b</sup>	No. of nonsynonymous SNPs	Putative function
CTA0447(CT412)	975	3	2	<i>pmpA</i>
CTA0884(CT812)	1,531	7	3	<i>pmpD</i>
CTA0954(CT874)	878	12	9	<i>pmpI</i>
CTA0951(CT871)	1,013	15	6	<i>pmpG</i>
CTA0715(CT658)	335	17	3	<i>sfhB</i>
CTA0429(CT394)	392	18	10	<i>hrcA</i>
CTA0448(CT413)	1,751	19	13	<i>pmpB</i>
CTA0166(CT157)	403	20	15	<i>pld</i>
CTA0751(CT690)	321	21	1	<i>dppD</i>
CTA0747(CT686)	395	22	11	ABC transporter-associated protein
CTA0738(CT678)	245	24	3	<i>pyrH</i>
CTA0449(CT414)	1,784	28	20	<i>pmpC</i>
CTA0498(CT456)	1,106	33	25	Tarp
CTA0733(CT675)	356	33	5	<i>karG</i>
CTA0675(CT622)	651	39	20	CHLPN 76-kDa homolog
CTA0153(CT144)	285	42	14	Hypothetical protein
CTA0156(CT147)	1,449	47	30	Hypothetical protein
CTA0739(CT679)	282	58	13	<i>tsf</i>
CTA0053(CT049)	490	82	39	Hypothetical protein
CTA0054(CT050)	582	100	50	Hypothetical protein
CTA0949(CT869)	962	129	50	<i>pmpE</i>
CTA0055(CT05)	520	130	63	Hypothetical protein
CTA0952(CT872)	1,018	153	39	<i>pmpH</i>
CTA0742(CT681)	396	180	53	<i>ompA</i> (MOMP)
CTA0950(CT870)	1,034	260	84	<i>pmpF</i>

<sup>a</sup> The A/HAR-13 ORF is listed first, with the D/UW-3 homolog listed in parentheses.

<sup>b</sup> Total number of SNPs, both synonymous and nonsynonymous, found in the ORF of interest.

proteins (46). Autotransporter proteins contain three primary domains, (i) an amino-terminal signal sequence and (ii) a carboxy-terminal, transmembrane domain (translocation unit) that is responsible for the secretion of (iii) the passenger domain (located at the amino-terminal end of the mature protein) to the cell surface (16). The passenger domain can either remain tethered to the cell surface, be proteolytically cleaved and released into the cell's environment, or be proteolytically cleaved but remain closely associated with the bacterial surface (17). Sequence analysis of a subset of the *pmp* genes, in particular *pmpH*, has demonstrated conservation of sequence identity between members within a pathogenic subgroup but a high degree of divergence between the different biovars (41). The above-mentioned results imply a role in pathogenic determination for at least a subset of the *pmp* genes. In addition, anti-PmpD antibodies neutralize infectivity by blocking entry (46), suggesting a key role for this protein, and perhaps related members of the gene family, in the pathogenesis of early infection.

While *pmpF* accounts for 0.3% (3,102/1,051,969) of the genomic sequence, it contains 8.3% (260/3,121) of all SNPs associated with ORFs, resulting in a total of 84 AAS. This represents a remarkable degree of targeted divergence be-

tween the two genomes. Therefore, we subjected PmpF to the SYFPEITHI algorithm (33) to determine if there was a correlation between AAS clustering and predicted MHC epitopes. Our analysis showed that the majority of the AAS are clustered within the central region of the passenger domain, coinciding with predicted epitope binding sites of different HLA class I and II alleles. Vandahl et al. (45) recently demonstrated that the amino-terminal half of a predicted autotransporter protein can be detected in the cytosol of *C. pneumoniae*-infected host cells. Given that other Pmps have been associated with the chlamydial cell surface (29, 44) and the fact that the PmpD ortholog of *C. pneumoniae* is processed in a manner consistent with its being an autotransporter (46), we believe the PmpF passenger domain is likely transported to the host cell cytosol. In addition, it is well established that foreign cytosolic proteins are susceptible to the endogenous antigen-presenting pathway and therefore become antigenic targets for CD8<sup>+</sup> cytotoxic T cells. Taken together, these findings suggest that PmpF could be a target of CD8<sup>+</sup> cytotoxic T cells. In addition, the passenger domain also has a high number of predicted class II epitopes, potentially implicating PmpF as a target for cytotoxic CD4<sup>+</sup> T cells, an effector phenotype that has recently been described for other microbial pathogens (7). Whether cytotoxic CD4<sup>+</sup> T cells function in chlamydial immunity is unknown. Class II is highly expressed in epithelial cells following exposure to gamma interferon (IFN- $\gamma$ ), a cytokine important for immunity to chlamydial infection (28). Moreover, IFN- $\gamma$  induces persistent aberrant chlamydial forms in infected epithelial cells (2). In models of penicillin-induced persistence, chlamydia-laden vesicles have been observed via electron microscopy blebbing from the inclusion membrane into the cytosol (D. K. Giles, J. D. Whitmore, R. W. LaRue, J. E. Raulston, and P. B. Wyrick, Abstr. 104th Gen. Meet. Am. Soc. Microbiol., abstr. D-221, 2004). Collectively these findings indicate that a chlamydial antigen(s) might be intercepting the class II exogenous antigen-processing pathway in IFN- $\gamma$ -exposed, chlamydia-infected epithelial cells, presenting a scenario for class II antigen presentation on the epithelial surface that provides a target for cytotoxic CD4<sup>+</sup> T cells. Studies focused on a Pmp antigen(s) as a target of chlamydial cytotoxic T cells seem warranted based on our observations. If proven to be the case, our SNP analysis indicates that mutations in class I and II antigenic sites for PmpF reflect a possible strategy for immune evasion by the pathogen.

It would be interesting to determine if *ompA* and *pmpF* variations are coupled or occur independently of one another. This question could be investigated by sequencing the respective genes from plaque-cloned EB populations of clinical isolates. If *pmpF* mutations occur independently of *ompA* mutations in clonal populations, it would implicate *pmpF* as an important target of protective immunity. In addition, if *pmpF* mutations are occurring to evade host immunity, this could explain why natural immunity and vaccine-induced immunity are short-lived in chlamydial infections. Partial, but incomplete, protection may be conferred by OmpA immunity, while complete immunity may require PmpF or other Pmp-specific host responses. Pmp antigenic variation may be an attempt by the chlamydiae to circumvent a protective response.

While the divergence of *ompA* and the *pmp* genes is of obvious historical and current relevance, perhaps as important



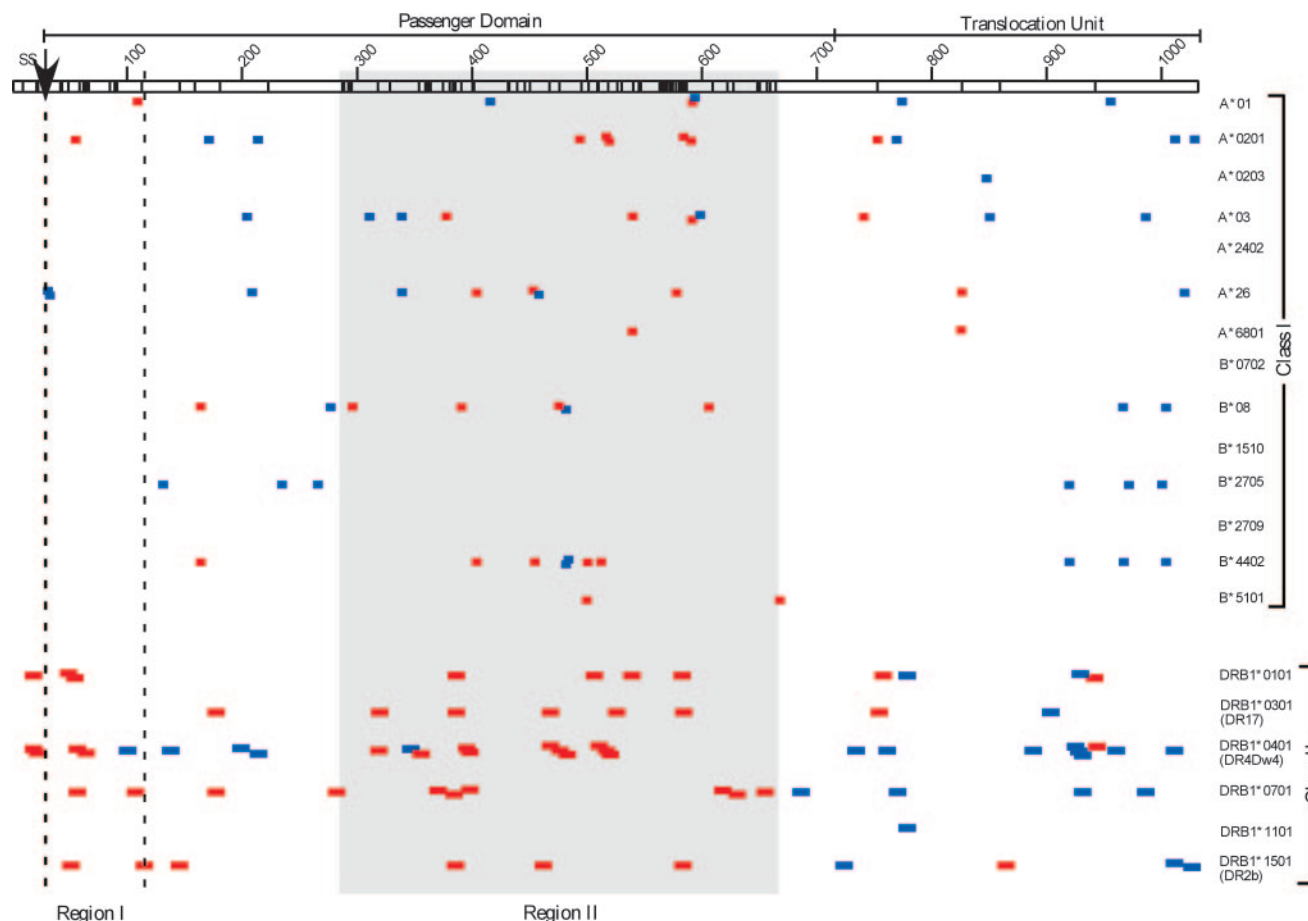


FIG. 5. MHC class I and II haplotype epitope mapping of PmpFs from both A/HAR-13 and D/UW-3. PmpF is indicated as an open box (top) and is 1,034 aa in size in both strains. The protein is predicted to be a member of a family of autotransporter proteins (15). As such, the predicted passenger domain, translocation unit, and site of signal sequence cleavage (SS ↓) are indicated above the open box. The PmpF sequences from both strains were aligned, and the location of each amino acid substitution is indicated with a vertical line within the open box. Two regions were identified via this alignment as containing a disproportionate number of the amino acid substitutions. Region I (87 aa) is located within the vertical dashed lines, while region II (378 aa) is located within the shaded region. The sequence of each protein was analyzed with the algorithm SYFPEITHI (33) to predict the locations of the MHC class I and II epitopes. The haplotypes included in this analysis are indicated on the right, with predicted epitopes identified for each haplotype located in the same row. Only epitopes giving a predictive score of 25 or greater are shown. Class I epitopes ranged in size from 8 to 10 aa, while class II epitopes were 15 aa in size. Epitopes that were found within conserved regions of the protein are indicated in blue. Epitopes found in regions shown to contain amino acid substitutions are indicated in red.

is a group of hypothetical serovar A genes that also exhibit a high degree of differentiation. Of particular interest is the disproportionate or higher percentage of SNPs observed in CTA0053 to CTA0055 compared to *pmpF* or *ompA*. While the functions of these genes remain unknown, the high degree of sequence divergence suggests that the proteins are immunogenic and are undergoing a high rate of mutation that may be involved in immune avoidance.

Tarp is rapidly tyrosine phosphorylated at the site of host cell entry and is implicated in actin recruitment (10). As noted by Clifton et al. (10), the 5' repeat region of L2/LGV-434 has six repeat elements, each containing four or five tyrosine residues. L2/LGV-434 Tarp has a total of 26 tyrosine residues in this repeat region, while D/UW-3 has 14 tyrosines and A/HAR-13 contains 13 tyrosines. Clifton et al. (9) have recently demonstrated that this region is the site of tyrosine phosphorylation. We have shown that the CT456 orthologs contain varying numbers of potential phosphorylation sites be-

tween the different serovars. The LGV strains contain the highest number of putative phosphorylation sites among *C. trachomatis* serovars, clearly differentiating invasive strains from noninvasive strains. One can hypothesize that the LGV Tarps may undergo an increased level of phosphorylation during early infection periods that intensifies their entry and/or actin nucleation properties at rates greater than those achieved by noninvasive serovars. How this might correlate with differences in host infection tropism (macrophages versus epithelial cells) and invasive versus noninvasive pathogenic characteristics is not intuitively obvious. Since Tarp is exposed to the cytoplasm of the host cell (10), the simplest explanation is that the putative enhanced Tarp phosphorylation of disseminating LGV serotypes allows a more efficient entry process that in turn augments their ability to circumvent host defense mechanisms by more efficiently modifying the inclusion membrane to prevent fusion with host lysosomes. A putative functional role for the 3' repeats of Tarp is less clear, although the

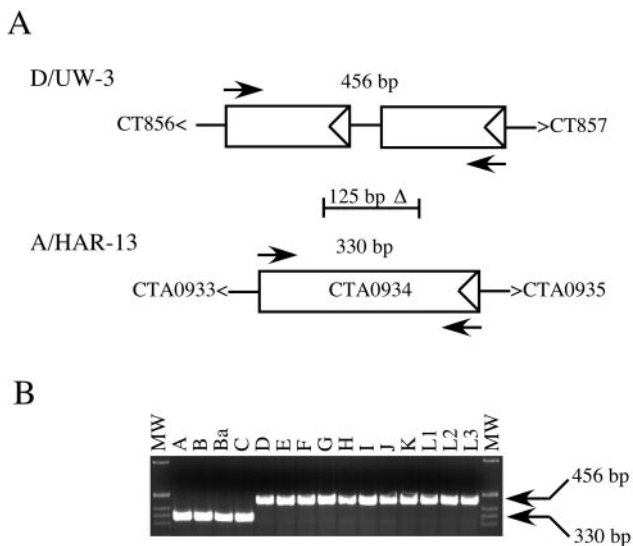


FIG. 6. Diagnostic PCR amplification identifying oculotropic versus genitotropic strains. (A) Comparison of the ORF arrangement for the genitotropic strain D/UW-3 versus that of the oculotropic strain A/HAR-13. The arrows indicate the relative primer binding sites for the diagnostic PCR amplification with the predicted base pair product size noted for each. The location of the 125-bp ocular-specific deletion is also indicated relative to D/UW-3. In addition to the 125-bp deletion, there is a single-nucleotide deletion located in CTA0934, accounting for the 126-bp difference in PCR product size between the two strains. (B) PCR amplification of the 15 *C. trachomatis* reference serovars. The sizes of the 1-kb DNA ladder (Invitrogen) bands shown are 1,018, 517/506, 396, 344, and 298 bp (from top to bottom).

C-terminal end of Tarp was recently identified as the region responsible for actin recruitment (9). Whether the C-terminal repeat elements are responsible for actin recruitment is not known. In addition, the fact that the serovar H ortholog exhibits the same number of 3' repeat elements as the LGV serovars argues against its being the sole determinant for the invasive phenotype. Finally, the LGV-specific repeats of six or more associated with the 5' portion of the gene should provide a genetic marker for differentiating between the invasive and noninvasive serovars. Further analysis employing more clinical isolates is required to verify this conclusion.

It is clear from all chlamydial genome-sequencing projects done to date that chromosomal gene content and synteny are highly conserved (20, 34, 35, 39). With regard to potential phenotypic differences, it is noteworthy that the ability to metabolize tryptophan is highly variable between species, as well

as between strains in the same species (Table 5). The chlamydial *trp* phenotype is significant because of the connection between tryptophan availability, chlamydial persistence, and host immune function via the protective effects of IFN- $\gamma$ . In human epithelial cells in vitro, IFN- $\gamma$  induces the expression of indoleamine 2,3-dioxygenase, an enzyme that degrades intracellular tryptophan. In cell culture systems, indoleamine 2,3-dioxygenase-induced tryptophan starvation has been implicated in the establishment of persistent chlamydial infection (6, 12). There is wide variation in the sensitivities of the *C. trachomatis* serovars to the inhibitory effects of IFN- $\gamma$  (27). In general, oculotropic serovars are much more sensitive than the genitotropic serovars, suggesting that *trpBA* may play a role in differential IFN- $\gamma$  sensitivity. In addition, Gieffers et al. (14) reported a correlation between the numbers of tandemly repeated *tyrP* copies in different *C. pneumoniae* isolates and their tissue origins of isolation (i.e., vascular versus respiratory isolates). Moreover, the authors demonstrated a correlation between copy number and amino acid transport capacity. It is interesting that the second copy of the tandemly repeated *tyrP* gene in A/HAR-13, CTA0892, is disrupted while both *tyrP* genes remain intact in D/UW-3. Under tryptophan-limiting conditions, this may result in a more dramatic decrease in tryptophan transport by serovar A than D and could provide an additional explanation for the increased sensitivity to IFN- $\gamma$ . It is intriguing that the oculotropic serovar has inactivated both genes needed to biosynthesize and transport tryptophan. Taken together, these results imply that it may be more advantageous for ocular serovars to induce a state of persistence.

In summary, the remarkable sequence identity between two chlamydial strains that show clearly distinct tissue tropisms and disease pathologies emphasizes the importance a limited number of small genomic variations and SNPs can have on an organism's phenotype. One important practical outcome of our efforts was the discovery of a diagnostic marker to differentiate between genitotropic and oculotropic strains (CTA0934), as well as a potential target for differentiating between invasive and noninvasive serovars (Tarp). Identification of these markers allows the simultaneous identification and organotropism differentiation of a chlamydial infection, a process previously requiring two separate assays. These markers will be useful tools in studying and tracking chlamydial epidemiology in third-world countries, where both blinding trachoma and chlamydial STDs are endemic.

Finally, in light of the remarkable degree of identity demonstrated between these two genomes, it is likely unnecessary to perform classical (shotgun) sequencing analysis of other *C. trachomatis* reference or clinical strains. Although comparative genomic analysis of other strains is a worthy scientific goal, it can be accomplished effectively by whole-genome DNA tiling microarrays (26). This technology is particularly suited for defining subtle differences (i.e., SNP or insertion/deletion) among related genomes. Moreover, DNA tiling microarray is significantly less costly and time-consuming, making it particularly attractive for high-throughput genomic comparisons of multiple *C. trachomatis* genomes.

#### ACKNOWLEDGMENTS

We are grateful for the technical contributions of Bill Whitmire and Debbie Crane. We thank Theresa Walunas and John Campbell for the

TABLE 5. Summary of genes involved in tryptophan metabolism for several chlamydiales

Chlamydia	Tryptophan transport	Tryptophan biosynthesis
<i>C. psittaci</i> GPIC	1 intact <i>tyrP</i>	<i>trpRDCFBA</i>
<i>C. trachomatis</i> D/UW-3	2 intact <i>tyrP</i>	<i>trpRBA</i>
<i>C. trachomatis</i> A/HAR-13	1 intact <i>tyrP</i>	Nonfunctional <i>trpBA</i>
<i>C. muridarum</i>	2 intact <i>tyrP</i>	None
<i>C. pneumoniae</i>	1–3 intact <i>tyrP</i>	None
<i>C. abortus</i>	1 intact <i>tyrP</i>	None
Parachlamydia UWE25	4 <i>tyrP</i>	None

sequence assembly, preliminary annotation, and SNP analysis studies. We thank Ted Hackstadt, Bob Heinzen, David Nelson, Paul Beare, and Laszlo Kari for critical comments and Kelly Matteson for secretarial assistance in the preparation of the manuscript.

## REFERENCES

- Baehr, W., Y. X. Zhang, T. Joseph, H. Su, F. E. Nano, K. D. Everett, and H. D. Caldwell. 1988. Mapping antigenic domains expressed by *Chlamydia trachomatis* major outer membrane protein genes. Proc. Natl. Acad. Sci. USA 85:4000–4004.
- Beatty, W. L., G. I. Byrne, and R. P. Morrison. 1993. Morphologic and antigenic characterization of interferon gamma-mediated persistent *Chlamydia trachomatis* infection in vitro. Proc. Natl. Acad. Sci. USA 90:3998–4002.
- Brunelle, B. W., T. L. Nicholson, and R. S. Stephens. 2004. Microarray-based genomic surveying of gene polymorphisms in *Chlamydia trachomatis*. Genome Biol. 5:R42.
- Brunham, R. C., and J. Rey-Ladino. 2005. Immunology of chlamydia infection: implications for a *Chlamydia trachomatis* vaccine. Nature 5:149–161.
- Caldwell, H. D., J. Kromhout, and J. Schachter. 1981. Purification and partial characterization of the major outer membrane protein of *Chlamydia trachomatis*. Infect. Immun. 31:1161–1176.
- Caldwell, H. D., H. Wood, D. Crane, R. Bailey, R. B. Jones, D. Mabey, I. Maclean, Z. Mohammed, R. Peeling, C. Roshick, J. Schachter, A. W. Solomon, W. E. Stamm, R. J. Suchland, L. Taylor, S. K. West, T. C. Quinn, R. J. Belland, and G. McClarty. 2003. Polymorphisms in *Chlamydia trachomatis* tryptophan synthase genes differentiate between genital and ocular isolates. J. Clin. Invest. 111:1757–1769.
- Canaday, D. H., R. J. Wilkinson, Q. Li, C. V. Harding, R. F. Silver, and W. H. Boom. 2001. CD4+ and CD8+ cells kill intracellular *Mycobacterium tuberculosis* by a perforin and Fas/Fas ligand-independent mechanism. J. Immunol. 167:2734–2742.
- Carlson, J. H., S. Hughes, D. Hogan, G. Cieplak, D. Sturdevant, G. McClarty, H. D. Caldwell, and R. J. Belland. 2004. Polymorphisms in the *Chlamydia trachomatis* cytotoxin locus associated with ocular and genital isolates. Infect. Immun. 72:7063–7072.
- Clifton, D. R., C. A. Dooley, S. S. Grieshaber, R. A. Carabeo, K. A. Fields, and T. Hackstadt. 2005. Tyrosine phosphorylation of chlamydial Tarp is species specific and not required for the recruitment of actin. Infect. Immun. 73:3860–3868.
- Clifton, D. R., K. A. Fields, S. Grieshaber, C. A. Dooley, E. Fischer, D. Mead, R. A. Carabeo, and T. Hackstadt. 2004. A chlamydial type III translocated protein is tyrosine phosphorylated at the site of entry and associated with recruitment of actin. Proc. Natl. Acad. Sci. USA 101:10166–10171.
- DelVecchio, V. G., V. Kapatral, R. J. Redkar, G. Patra, C. Mujer, T. Los, N. Ivanova, I. Anderson, A. Bhattacharyya, A. Lykidis, G. Reznik, L. Jablonski, N. Larsen, M. D'Souza, A. Bernal, M. Mazur, E. Goltsman, E. Selkov, P. H. Elzer, S. Hagiya, D. O'Callaghan, J. J. Letesson, R. Haselkorn, N. Kyrpides, and R. Overbeek. 2002. The genome sequence of the facultative intracellular pathogen *Brucella melitensis*. Proc. Natl. Acad. Sci. USA 99:443–448.
- Fehlner-Gardiner, C., C. Roshick, J. H. Carlson, S. Hughes, R. J. Belland, H. D. Caldwell, and G. McClarty. 2002. Molecular basis defining human *Chlamydia trachomatis* tissue tropism. A possible role for tryptophan synthase. J. Biol. Chem. 277:26893–26903.
- Gardel, D., K. Johnson, A. Jacq, and J. Beckwith. 2005. The *secD* locus of *E. coli* for two membrane proteins required for protein export. EMBO J. 9:3209–3216.
- Gieffers, J., L. Durling, S. P. Ouellette, J. Rupp, M. Maass, G. I. Byrne, H. D. Caldwell, and R. J. Belland. 2003. Genotypic differences in the *Chlamydia pneumoniae* *tyrP* locus related to vascular tropism and pathogenicity. J. Infect. Dis. 188:1085–1093.
- Henderson, I. R., and A. C. Lam. 2001. Polymorphic proteins of *Chlamydia* spp.—autotransporters beyond the Proteobacteria. Trends Microbiol. 9:573–578.
- Henderson, I. R., and J. P. Nataro. 2001. Virulence functions of autotransporter proteins. Infect. Immun. 69:1231–1243.
- Henderson, I. R., F. Navarro-Garcia, and J. P. Nataro. 1998. The great escape: structure and function of the autotransporter proteins. Trends Microbiol. 6:370–378.
- Higgins, D. G., and P. M. Sharp. 1989. Fast and sensitive multiple sequence alignments on a microcomputer. Comput. Appl. Biosci. 5:151–153.
- Ivanova, N., A. Sorokin, I. Anderson, N. Galleron, B. Candelon, V. Kapatral, A. Bhattacharyya, G. Reznik, N. Mikhailova, A. Lapidus, L. Chu, M. Mazur, E. Goltsman, N. Larsen, M. D'Souza, T. Walunas, G. Grechkin, G. Pusch, R. Haselkorn, M. Fonstein, S. D. Ehrlich, R. Overbeek, and N. Kyrpides. 2003. Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. Nature 423:87–91.
- Kalman, S., W. Mitchell, R. Marathe, C. Lammel, J. Fan, R. W. Hyman, L. Olinger, J. Grimwood, R. W. Davis, and R. S. Stephens. 1999. Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*. Nat. Genet. 21:385–389.
- Kapatral, V., I. Anderson, N. Ivanova, G. Reznik, T. Los, A. Lykidis, A. Bhattacharyya, A. Bartman, W. Gardner, G. Grechkin, L. Zhu, O. Vasieva, L. Chu, Y. Kogan, O. Chaga, E. Goltsman, A. Bernal, N. Larsen, M. D'Souza, T. Walunas, G. Pusch, R. Haselkorn, M. Fonstein, N. Kyrpides, and R. Overbeek. 2002. Genome sequence and analysis of the oral bacterium *Fusobacterium nucleatum* strain ATCC 25586. J. Bacteriol. 184:2005–2018.
- Kim, S. K., M. Angevine, K. Demick, L. Ortiz, R. Rudersdorf, D. Watkins, and R. DeMars. 1999. Induction of HLA class I-restricted CD8+ CTLs specific for the major outer membrane protein of *Chlamydia trachomatis* in human genital tract infections. J. Immunol. 162:6855–6866.
- Kim, S. K., L. Devine, M. Angevine, R. DeMars, and P. B. Kavathas. 2000. Direct detection and magnetic isolation of *Chlamydia trachomatis* major outer membrane protein-specific CD8+ CTLs with HLA class I tetramers. J. Immunol. 165:7285–7295.
- Lindquist, E., and R. Stephens. 1998. Transcriptional activity of a sequence variable protein family in *Chlamydia trachomatis*, p. 259–262. In R. Stephens, G. Byrne, G. Christiansen, I. Clarke, J. Grayston, R. Rank, G. Ridgway, P. Saikku, J. Schachter, and W. Stamm (ed.), Proceedings of the Ninth International Symposium on Human Chlamydial Infection. International Chlamydia Symposium, San Francisco, Calif.
- Matsumoto, A., H. Izutsu, N. Miyashita, and M. Ohuchi. 1998. Plaque formation by and plaque cloning of *Chlamydia trachomatis* biovar trachoma. J. Clin. Microbiol. 36:3013–3019.
- Mockler, T. C., and J. R. Ecker. 2005. Applications of DNA tiling arrays for whole-genome analysis. Genomics 85:1–15.
- Morrison, R. P. 2000. Differential sensitivities of *Chlamydia trachomatis* strains to inhibitory effects of gamma interferon. Infect. Immun. 68:6038–6040.
- Morrison, R. P., and H. D. Caldwell. 2002. Immunity to murine chlamydial genital infection. Infect. Immun. 70:2741–2751.
- Myygind, P. H., G. Christiansen, P. Roepstorff, and S. Birkelund. 2000. Membrane proteins PmpG and PmpH are major constituents of *Chlamydia trachomatis* L2 outer membrane complex. FEMS Microbiol. Lett. 15:163–169.
- Ortiz, L., M. Angevine, S. K. Kim, D. Watkins, and R. DeMars. 2000. T-cell epitopes in variable segments of *Chlamydia trachomatis* major outer membrane protein elicit serovar-specific immune responses in infected humans. Infect. Immun. 68:1719–1723.
- Ortiz, L., K. P. Demick, J. W. Petersen, M. Polka, R. A. Rudersdorf, B. Van der Pol, R. Jones, M. Angevine, and R. DeMars. 1996. *Chlamydia trachomatis* major outer membrane protein (MOMP) epitopes that activate HLA class II-restricted T cells from infected humans. J. Immunol. 157:4554–4567.
- Overbeek, R., N. Larsen, T. Walunas, M. D'Souza, G. Pusch, E. Selkov, K. Liolios, V. Joukov, D. Kaznadzey, I. Anderson, A. Bhattacharyya, H. Burd, W. Gardner, P. Hanke, V. Kapatral, N. Mikhailova, O. Vasieva, A. Osterman, V. Vonstein, M. Fonstein, N. Ivanova, and N. Kyrpides. 2003. The ERGO genome analysis and discovery system. Nucleic Acids Res. 31:164–171.
- Rammensee, H. G., J. Bachmann, N. P. Emmerich, O. A. Bachor, and S. Stevanovic. 1999. SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics 50:213–219.
- Read, T. D., R. C. Brunham, C. Shen, S. R. Gill, J. F. Heidelberg, O. White, E. K. Hickey, J. Peterson, T. Utterback, K. Berry, S. Bass, K. Linher, J. Weidman, H. Khouri, B. Craven, C. Bowman, R. Dodson, M. Gwinn, W. Nelson, R. DeBoy, J. Kolonay, G. McClarty, S. L. Salzberg, J. Eisen, and C. M. Fraser. 2000. Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. Nucleic Acids Res. 28:1397–1406.
- Read, T. D., G. S. Muers, R. C. Brunham, W. C. Nelson, I. T. Paulsen, J. F. Heidelberg, E. Holtzapple, H. Khouri, N. B. Federova, H. A. Carty, L. A. Umayam, D. H. Haft, J. Peterson, M. J. Beanan, O. White, S. L. Salzberg, R. Hsia, G. McClarty, R. G. Rank, P. M. Bavoil, and C. M. Fraser. 2003. Genome sequence of *Chlamydiaophila caviae* (*Chlamydia psittaci* GPIC): examining the role of niche-specific genes in the evolution of the Chlamydiae. Nucleic Acids Res. 31:2134–2147.
- Schachter, J. 1978. Chlamydial infections (first of three parts). N. Engl. J. Med. 298:428–434.
- Schachter, J. 1978. Chlamydial infections (second of three parts). N. Engl. J. Med. 298:490–495.
- Schachter, J. 1978. Chlamydial infections (third of three parts). N. Engl. J. Med. 298:540–549.
- Stephens, R. S., S. Kalman, C. Lammel, J. Fan, R. Marathe, L. Aravind, W. Mitchell, L. Olinger, R. L. Tatusov, Q. Zhao, E. V. Koonin, and R. W. Davis. 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. Science 282:754–759.
- Stephens, R. S., M. R. Tam, C. C. Kuo, and R. C. Nowinski. 1982. Monoclonal antibodies to *Chlamydia trachomatis*: antibody specificities and antigen characterization. J. Immunol. 128:1083–1089.
- Stothard, D. R., G. A. Toth, and B. E. Batteiger. 2003. Polymorphic membrane protein H has evolved in parallel with the three disease-causing groups of *Chlamydia trachomatis*. Infect. Immun. 71:1200–1208.



42. **Su, H., and H. D. Caldwell.** 1991. In vitro neutralization of *Chlamydia trachomatis* by monovalent Fab antibody specific to the major outer membrane protein. *Infect. Immun.* **59**:2843–2845.
43. **Su, H., N. G. Watkins, Y. X. Zhang, and H. D. Caldwell.** 1990. *Chlamydia trachomatis*-host cell interactions: role of the chlamydial major outer membrane protein as an adhesin. *Infect. Immun.* **58**:1017–1025.
44. **Tanzer, R. J., and T. P. Hatch.** 2001. Characterization of outer membrane proteins in *Chlamydia trachomatis* LGV serovar L2. *J. Bacteriol.* **183**:2686–2690.
45. **Vandahl, B. B., A. Stensballe, P. Roepstorff, G. Christiansen, and S. Birke-lund.** 2005. Secretion of Cpn0796 from *Chlamydia pneumoniae* into the host cell cytoplasm by an autotransporter mechanism. *Cell Microbiol.* **6**:825–836.
46. **Wehrli, W., V. Brinkmann, P. R. Jungblut, T. F. Meyer, and A. J. Szczepek.** 2004. From the inside out—processing of the chlamydial autotransporter PmpD and its role in bacterial adhesion and activation of human host cells. *Mol. Microbiol.* **51**:319–334.
47. **Yuan, Y., Y.-X. Zhang, N. G. Watkins, and H. D. Caldwell.** 1989. Nucleotide and deduced amino acid sequences for the four variable domains of the major outer membrane proteins of the 15 *Chlamydia trachomatis* serovars. *Infect. Immun.* **57**:1040–1049.
48. **Zhang, Y. X., S. J. Stewart, and H. D. Caldwell.** 1989. Protective monoclonal antibodies to *Chlamydia trachomatis* serovar- and serogroup-specific major outer membrane protein determinants. *Infect. Immun.* **57**:636–638.

---

Editor: J. B. Bliska