Molecular and Mutation Trends Analyses of *omp1* Alleles for Serovar E of *Chlamydia trachomatis*

Implications for the Immunopathogenesis of Disease

Deborah Dean*‡ and Kim Millman‡

*Department of Medicine, *Division of Infectious Diseases, and ‡The Francis I. Proctor Foundation, University of California San Francisco School of Medicine, San Francisco, California 94143-0412*

Abstract

Serovars E, F, and D are the most prevalent *Chlamydia trachomatis* strains worldwide. This prevalence may relate to epitopes that enhance infectivity and transmission. There are numerous major outer membrane protein (MOMP) gene (*omp1*) variants described for D and F but few for E. However, *omp1* constant regions are rarely sequenced, yet, they may contain mutations that affect the structure/function relationship of the protein. Further, differentiating variants that occur as a result of selection from variants that contain random mutations without biologic impact is difficult. We investigated 67 urogenital E serovars and found 11 (16%) variants which contained 16 (53%) nonconservative amino acid changes. Using signature-pattern analysis, 57 amino acids throughout MOMP differentiated the E sequence set from the non-E sequence set, thus defining E strains. Four E variants did not match this signature-pattern, and, by phenetic analyses, formed new phylogenetic branches, suggesting that they may be biologically distinct variants. Our analyses offer for the first time a unique approach for identifying variants that may occur from selection and may affect infectivity and transmission. Understanding the mutation trends, phylogeny, and molecular epidemiology of E variants is essential for designing public health control interventions and a vaccine. (*J. Clin. Invest.* 1997. 99:475–483.) Key words: *Chlamydia trachomatis* • serovar E alleles • signature-pattern analysis • antigenic variation • *omp1* genotypes

Introduction

The predominant *Chlamydia trachomatis* strains that are responsible for sexually transmitted diseases (STD)[1] worldwide are serovars E, F, and D (1–5). The public health costs for the morbidity and sequelae of these infections have reached $2.4 billion per year (6). Thus, there is an urgent need to develop appropriate public health interventions and pursue a vaccine for chlamydial STDs.

A major focus for chlamydial vaccine research is the major outer membrane protein (MOMP) which is the most antigenically diverse and abundant surface protein of the organism. MOMP has become an important target for epitope mapping to define surface exposed regions for antibody binding and neutralization studies, for determining host cell–mediated immune interactions, and for immunotyping to identify the spectrum of strains responsible for human disease. The mAbs used for serotyping recognize serovar-specific determinants of MOMP (7) which reflect many but not all of the amino acid variations found among the 18 known serovars of *C. trachomatis.* These are located within three of the four variable sequence regions of the protein, termed variable segments (VS) 1, 2, 3, and 4 (8, 9). Epitope mapping has revealed convincing evidence that VS1, 2, and 4 are surface exposed. In contrast, VS3 does not appear to be surface exposed but can elicit T cell help for antibody production directed at MOMP epitopes (10). Thus, MOMP displays an array of molecular biologic and immunologic characteristics that are important for the study of this pathogen. Further, understanding the diversity of MOMP at the nucleotide and protein level among serovars will be important for vaccine design.

Towards this goal, a number of investigators have examined the sequence diversity of the MOMP gene (*omp1*). This research has revealed an array of genotypic variation within each of the known serovars (1, 2, 11–17). Many of these variants contain amino acid changes which cannot be identified by serotyping (18, 19). Among STD populations, a relatively high number of variants have been described among D and F serovars, yet relatively few have been identified for serovar E (1, 2, 11–13). This may reflect the fact that only VS1, 2, and 4 sequences have been used for *omp1* genotyping, not VS3 or constant regions. In a study by Frost et al. (13), nucleotide sequence variation in constant regions was able to differentiate ocular (trachoma) from genital samples of the same serotype. It is possible that the constant regions of E serovars may also contain sequence variation. This is important since sequence variation in both variable and constant regions may alter the structure and function of MOMP.

There is consensus in the field that *omp1* represents a greater degree of polymorphism than previously appreciated and there is speculation that strain variation may occur in response to host immune surveillance as *omp1* is the predominant antigen-specifying gene for *C. trachomatis* (2, 12, 17, 20). However, to date, there has been no approach for differentiating *omp1* variants that arise from random mutations versus those that occur in response to selective pressure. Since serovar E is the most prevalent of the chlamydial STD pathogens, we investigated these issues among serovar E and E variants.

Address correspondence to Dr. Deborah Dean, University of California San Francisco School of Medicine, Proctor Foundation, Box 0412, 513 Parnassus Ave., San Francisco, CA 94143-0412. Phone: 415-476-4548; FAX: 415-476-6085; E-mail: debd@itsa.ucsf.edu

Received for publication 15 August 1996 and accepted in revised form 22 November 1996.

1. *Abbreviations used in this paper:* MOMP, major outer membrane protein; *omp1*, major outer membrane protein gene; STD, sexually transmitted diseases; VS, variable segment.

J. Clin. Invest.

0021-9738/97/02/0475/09 $2.00
Volume 99, Number 3, February 1997, 475–483

67 urogenital specimens that immunotyped as serovar E were analyzed. These data provide for the first time the complete *omp1* sequence of these strains. In addition, we conducted phylogenetic analyses of the observed variants and prototype strains in the data set. We also used signature-pattern analysis to define an E strain in distinction from prototype non-E strains and to differentiate E variants that may be biologically significant.

Methods

Chlamydial strains. Reference strains E/Bour and E/UW-5/Cx were used for comparative immunotyping. Both strains were also included in the alignments and analyses of the E serovars identified in this study. In addition, prototype genital strains B/TW-5/OT, C/TW-3/OT, D/UW-3/Cx, F/UW-13/Ur, G/UW-57/Cx, H/UW-4/Cx, I/UW-12/Ur, J/UW-36/Cx, K/UW-31/Cx, LGV/L1-440, LGV/L2-434, and LGV/L3-404 were sequenced to provide the entire sequence of *omp1*.

Cervical and urethral samples were collected from women and men, respectively. These patients attended STD, obstetrics, and gynecology clinics in the San Francisco/San Jose Bay Area over a period of 4 yr from 1989 to 1992. None of the patients had been on antibiotics for at least 2 wk before examination and sample collection.

Isolation and immunotyping of chlamydial strains. All samples were stored at −80°C. The method of Ripa et al. (21) was used for propagation of *C. trachomatis* clinical samples and prototype serovars D, E, L1, and L2 in paired, 12-mm shell vials. This technique was modified slightly. Briefly, 200–400 μl of sample was inoculated onto McCoy cell monolayers, and one vial was stained after 48 h of growth with MOMP-specific fluorescein-conjugated mAbs (MicroTrak; Syva Co., Palo Alto, CA) as described previously (2). If negative, the paired vial was passaged once and stained after an additional 48 h. If either vial was positive, the organism was propagated to a level of ∼ 70% infectivity per monolayer. Immunotyping was performed according to the method of Wang et al. (7). Briefly, monolayers were sonicated, centrifuged, and 16 mAbs (Washington Research Foundation, Seattle, WA) were used to distinguish the serotype of each isolate according to the foundation's instructions.

Amplification of chlamydial DNA. Each of the prototype genital serovars were *omp1* genotyped. Samples that were positive by culture and immunotyped as serovar E were also subjected to *omp1* genotyping. Chlamydial DNA was extracted as described previously (14) from a 100-μl aliquot of nonpropagated sample in transport media that had been stored at −80°C. The DNA template was amplified by PCR using primers FII and BII which are specific for the MOMP gene (2). The 100-μl reaction mixture and thermocycler profile have been described previously (2). 1 μl of the first reaction was used in a nested PCR with primers MF21 and MB4 which flank VS1 and VS4 of *omp1* and are internal to FII and BII (22). To sequence the segment of *omp1* upstream of MF21, primers MF100 (5′ TGTAAAACGACGGCCAGTGCTTCCCTCCTTGCAAGCTCT 3′, bp 21–41) and B100 (5′ GGCACCCATTTGGAATTCTTT 3′, bp 237–216) were designed to amplify this region. The reagents and thermocycling profile were the same as described previously (2). 10 μl of each reaction product was run on a 1.5% agarose gel with ethidium bromide and molecular weight markers to confirm the size of the amplified DNA. Only the amplified DNA from the nested and upstream reactions were sequenced after purification of the DNA from agarose gels using GeneClean II (BIO 101, Inc., Vista, CA).

Cloning methods for PCR products. The PCR products resulting from a separate amplification of the original sample using primers F11 and B11 were cloned using the TA cloning system (Invitrogen, San Diego, CA) according to the manufacturer's instructions. The ligation with the pCR™ II vector was accomplished with a 1:3 molar ratio of vector to PCR insert. Transformation of competent INVαF′ *Escherichia coli* cells was performed using standard techniques and plated on LB plates containing 50 μg/ml of ampicillin, and incubated overnight at 37°C. The DNA was maxiprepped, purified by GeneClean II, and sequenced as described below.

omp1 genotyping. The chlamydial DNA from both PCR and cloned samples were subjected to *omp1* genotyping which was performed as described previously (2). Briefly, ∼ 60 ng of each template was annealed to M13 universal dye primers, cycle sequenced by PCR, and electrophoresed on a standard gel in the 373A Automated Sequencing System (Applied Biosystems, Foster City, CA). Each sequence was read in a blinded fashion by automation and semiautomation.

Genotyping analysis. Each *omp1* sequence generated by the automated system was downloaded to the computer and verified by semiautomation as described previously (2). The sequence, identification number, and date of collection were then loaded onto GeneWorks (IntelliGenetics, Inc., Mountain View, CA) for sequence alignment and comparison between PCR and cloned results from the same samples, and for comparison with all other E genotype sequences. Each unique DNA sequence was operationally defined as a genotype. There were no discrepancies between sequences of the same sample. Sequences that differed within VS compared with the prototype E sequence were designated as a variant and as a single variable for data analysis. A two-tailed Fisher's Exact Test was used for comparison of proportions.

Computational analyses. The amino acid sequences of MOMP for each E strain in this study and prototype MOMP sequences for genotypes B, C, D, F, G, H, I, J, and K were aligned using PIMA (23). This program provides a statistical algorithm for analysis using a scoring matrix for gaps and amino acid side chain chemistry, and a progressive alignment strategy for aligning multiple sequences. After the amino acids were aligned, nucleotide alignments were generated based on amino acid alignments and nucleotide differences not observed at the amino acid level. These manipulations were accomplished using the multiple alignment sequence editor MASE (24). After alignment, a phylogenetic tree was constructed. Unweighted-parsimony was generated by PAUP (25) for analysis using MacClade (25). 1/f, where f is the frequency of character state changes calculated by MacClade, was used to weigh possible nucleotide changes in order to account for possible asymmetrical substitution frequencies of mutations. A weighted character type matrix was calculated and used by PAUP to generate weighted-parsimony. The most parsimonious tree is that which has the sum of the branch links minimized, and the branch links represent the number of weighted character changes. The relative branch lengths in the tree were drawn from 10 random inputs of the taxa order using the heuristic searches to generate the most parsimonious tree; the phylogenetic drawing tool in PAUP was used to display the tree. Bootstrap values were calculated for weighted-parsimony after 100 bootstrap replicates whereby the branch points on the tree represent the 50% threshold majority consensus values (see Fig. 4). PHYLIP (26), using Fitch-Margoliash and least-squares distance methods, was used to determine the phylogenetic relatedness of the sequence set.

The VESPA (27) program was used to identify atypical nucleotide or amino acids in a query set of sequences (all E prototype sequences and variants) in comparison with a reference sequence (all non-E genital strains). The frequency of the most common nucleotide or amino acid was calculated for each position in the reference sequence and the query sequence set. Three signature-patterns can be defined (27): (*1*) stringent sequences where the atypical amino acids from the sequence set have to be present in all sibling sequences included in the set; (*2*) majority sequences where the most common amino acid in the set differs from the most common in the reference sequence; and (*3*) direct sequences which are experimentally derived consensus sequences. Signature-patterns were also identified by comparing the entropy value for each sequence from the non-E genital sequences, comparing it with a set of E sequences within the sequence alignment. Entropy (H(i)) relies on probabilities and defines both the variety and frequency of observed amino acids at each position within the sequence (28). A signature-pattern describes amino

Table I. mAb Typing of Cervical Isolates Representing Prototype and Variant E Serovars Compared with Prototype D, L1, and L2 Immunotyping Patterns

Serotype specificity	mAb	Prototype E serovars	E/1	E/2	Prototype D serovar	Prototype L1 serovar	Prototype L2 serovar
Species-specific	LV-22	++++	++++	++++	++++	++++	++++
All B complex	BB-11	++++	++++	++++	++++	++++	++++
Partial B complex	LV-23	++++	++++	++++	−	−	++++
L2-specific	LV-27	−	−	−	−	−	++++
E and L1-specific	KB-8	++++	++++	++++	−	++++	−
D-specific	JG-9	−	++	++	++++	−	−

acids that are very common or identical in one set of sequences (e.g., non-E sequences) but differ among all the E genotypes. A Monte Carlo randomization was used to determine the statistical significance of difference in entropy. For a conserved signature position to be statistically significant, the positions in the randomized data sets did not have an entropy value as high as that found in the original data sets (28). This type of analysis can evaluate signature-pattern characteristics that might distinguish an E from other chlamydial genotypes and their variants; identify crucial amino acids required for protein stability; and predict biologic properties from primary structure (transmission from a different reservoir, infection in a different tissue, or tissue tropism).

Results

Immunotyping. All samples that were positive by culture were propagated and immunotyped. Of 160 culture positive samples collected over 4 yr, 67 (42%) typed as serovar E. Table I summarizes the immunotyping results for two variant E genotypes that displayed a unique typing pattern from the other variants. These two were compared with the typing patterns for prototype E serovars and closely related B complex prototype serovars: D, L1, and L2. B complex serovars are comprised of serovars B, Ba, D, E, L1, and L2 because each binds to subspecies-spe-

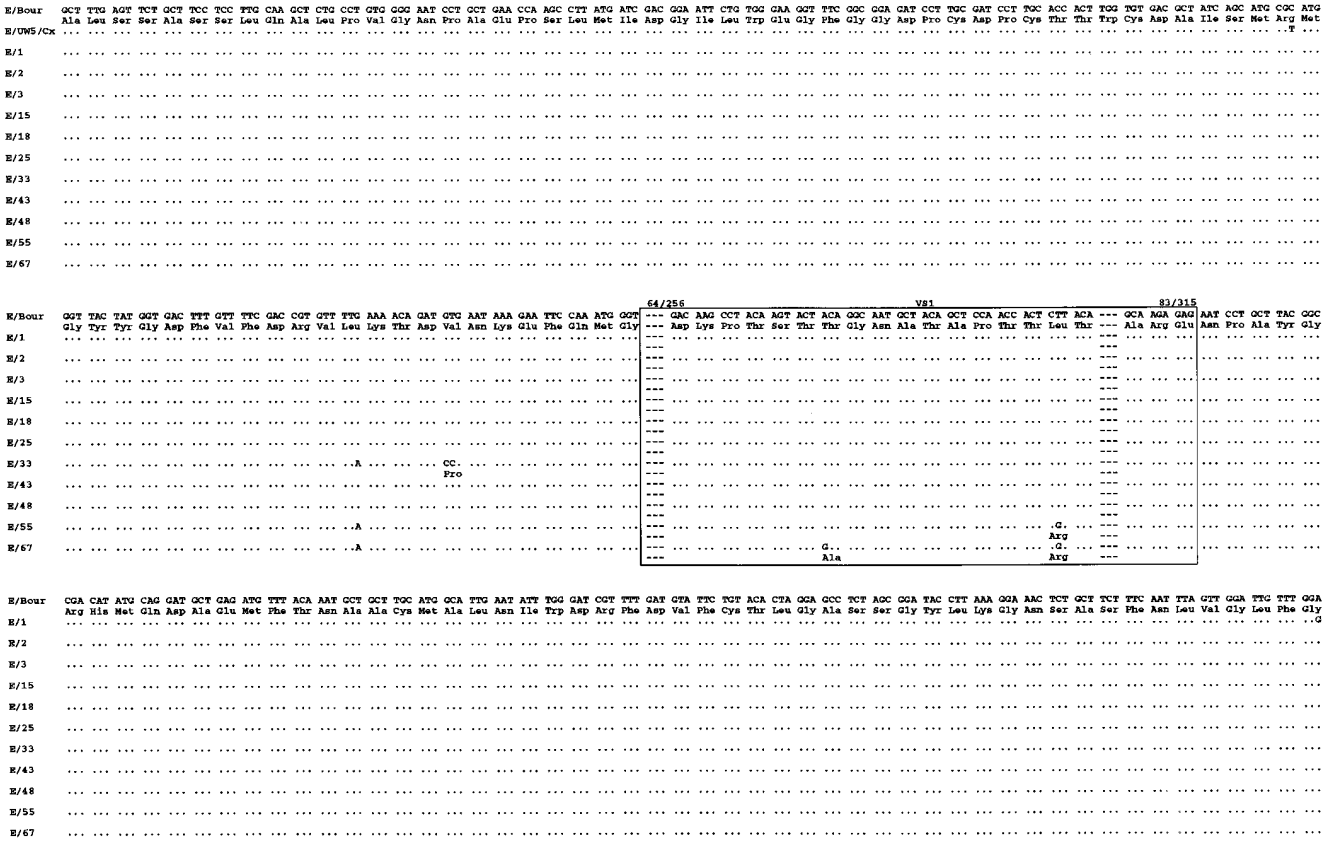

Figure 1. Nucleotide and deduced amino acid sequence of prototype strain E/Bour and variant E genotypes. The E variants are denoted by numbers after the backslash. The boxed region denotes VS1 of *C. trachomatis*. The numbers at the top corners of the box are the amino acid position followed by the nucleic acid position (after the backslash) on *omp1* according to Yuan et al. (36). The dots represent homology with the nucleotide sequence of E/Bour. The dashes represent a lack of a codon when aligned with other non-E serovars using the reference sequences of Yuan et al. (36). These sequence data are available under GenBank accession numbers U78528 to U78538 for sequences E/1 to E/67, respectively, and U78763 for E/Bour.

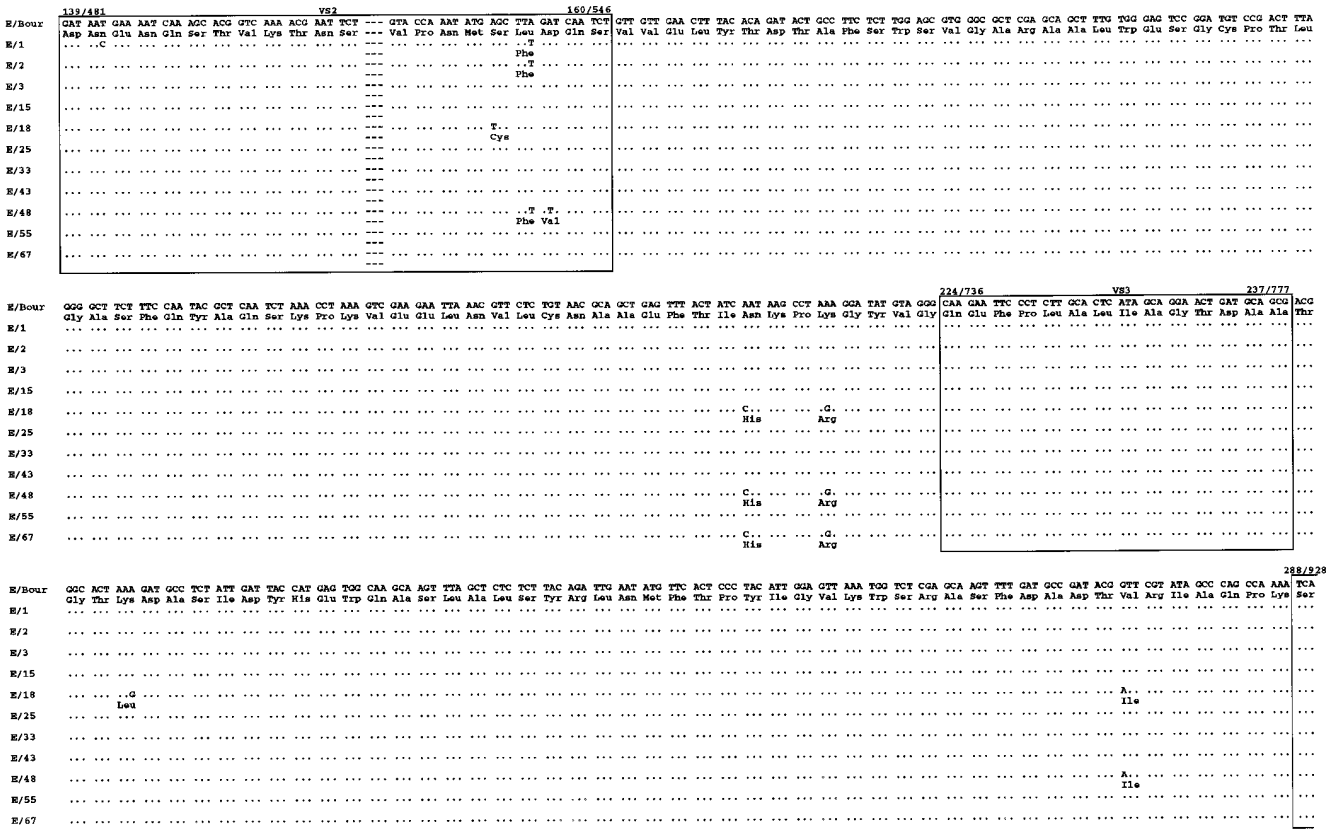

Figure 2. Nucleotide and deduced amino acid sequence of prototype strain E/Bour and variant E genotypes. The E variants are denoted by numbers after the backslash. Boxed regions denote VS2 and VS3 of *C. trachomatis*. The numbers at the top corners of the boxes are the amino acid position followed by the nucleic acid position (after the backslash) on *omp1* according to Yuan et al. (36). The dots represent homology with the nucleotide sequence of E/Bour. The dashes represent a lack of a codon when aligned with other non-E serovars using the reference sequences of Yuan et al. (36). These sequence data are available under GenBank accession numbers U78528 to U78538 for sequences E/1 to E/67, respectively, and U78763 for E/Bour.

cific mAb BB-11 unlike the other serovars. Samples E/1 and E/2 were the only variants to react with mAb JG-9 which is specific for serovar D (Table I). However, the degree of reactivity to JG-9 was only 2+ compared with 4+ for the other mAbs. Of note is the fact that both of these variants contained a single amino acid change in the carboxy-terminal region of VS2 which was conservative (see Fig. 2) and not expected to significantly alter the protein. These changes resulted in a sequence identical to that of prototype D sequences in this region. Epitope mapping to determine the exact binding site for JG-9 has not been performed. The lack of differentiation of E variants by immunotyping may reflect the limitations of current mAbs in that only certain epitopes can be recognized.

Correlation of immunotyping with genotyping analyses. Of the 67 samples immunotyped as serovar E, *omp1* genotyping revealed 11 (16%) E genotype variants compared with prototype E genotypes E/Bour and E/UW-5/Cx. E/UW-5/Cx differed from E/Bour by one nucleotide that was upstream of VS1 at nucleotide position 183 (Fig. 1). Figs. 1–3 show the nucleotide and presumed amino acid sequences of each variant in comparison with E/Bour and E/UW-5/Cx. Most variants were found in two or more individuals: two for E/2, E/33, E/43, and E/55, and three for E/1, E/18, E/48, and E/67. The sequence of one E genotype sample (sample 44) from Canada (1) is included in Fig. 3 for comparison with our variant sequences in VS4. This was the only E variant sequence found among the Winnipeg patient population and displayed one nucleotide and amino acid sequence change similar to the E/25 and E/43 variants described in this paper.

The 39 nucleotide substitutions resulted in 30 (77%) nonsynonymous codons and 7 (18%) synonymous codons; two codons contained two nucleotide substitutions each (Figs. 1–3). 25 (64%) of the 39 nucleotide substitutions for all variants were a purine transition, suggesting that these base substitutions were not due to misincorporation by *Taq* polymerase as pyrimidine substitutions are favored by this polymerase (29). In 8 (38%) of the 21 codon positions where nucleotide substitutions occurred, two or more variant E genotypes were found to have the same substitution. All of these eight substitutions resulted in nonsynonymous codons. There was almost equal distribution of nonsynonymous codons occurring within and outside of VS with the exception of VS3 where there were no nucleotide substitutions. However, there were significant nonconservative amino acid changes flanking this VS: Asp 216 His and Lys 241 Leu (Fig. 2). 16 (53%) of the 30 nonsynonymous codons were nonconservative amino acid changes; 12 (75%) of these 16 occurred in VS. Of note is the fact that nucleotide and amino acid changes at amino acid position 157 were identical

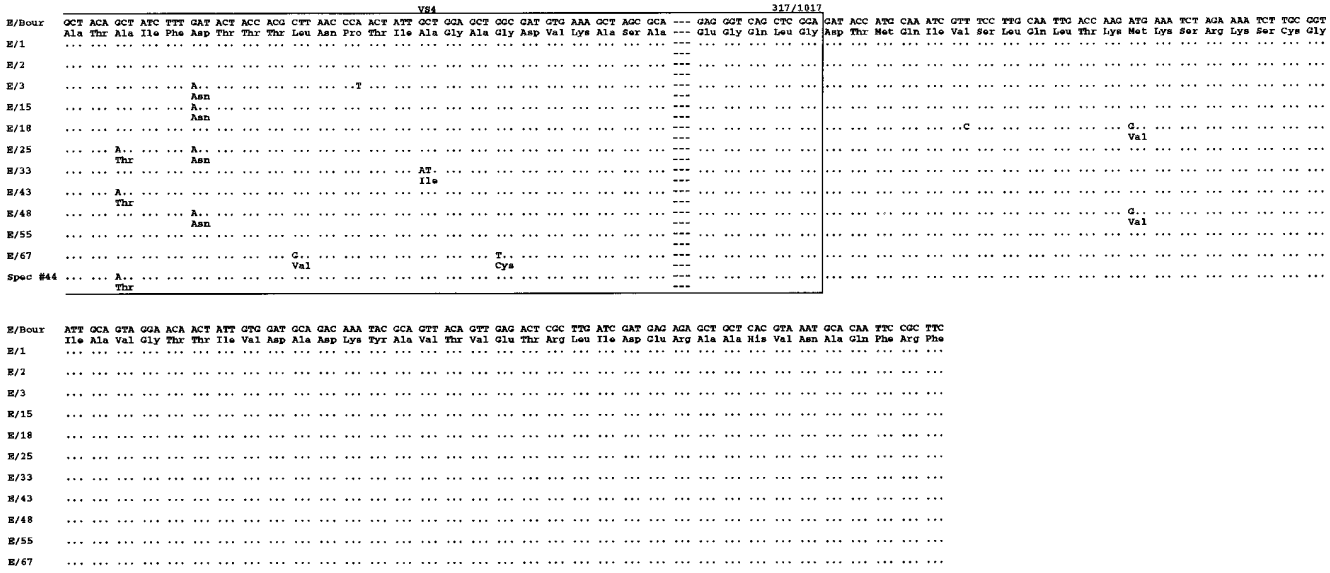

Figure 3. Nucleotide and deduced amino acid sequence of prototype strain E/Bour and variant E genotypes. The E variants are denoted by numbers after the backslash. The boxed region denotes VS4 of *C. trachomatis*. The numbers at the top corners of the box are the amino acid position followed by the nucleic acid position (after the backslash) on *omp1* according to Yuan et al. (36). The dots represent homology with the nucleotide sequence of E/Bour. The dashes represent a lack of a codon when aligned with other non-E serovars using the reference sequences of Yuan et al. (36). These sequence data are available under GenBank accession numbers U78528 to U78538 for sequences E/1 to E/67, respectively, and U78763 for E/Bour.

to prototype genotypes D and L1 (Fig. 2). These changes occurred in variants E/1 and E/2 where both reacted to D-specific MAb, JG-9 (Table I). Interestingly, variant E/48 also had this substitution. However, there was an adjacent nonconservative amino acid change: Asp 158 Val. This may have prevented the binding of JG-9 to this protein.

omp1 genotyping of cloned samples in comparison with PCR products. To determine whether PCR introduces nucleotide error in template reading which could produce variant *omp1* genotypes, samples were separately cloned and sequenced. Original sample material was used in a separate PCR to amplify *omp1* using different primers than for the PCR

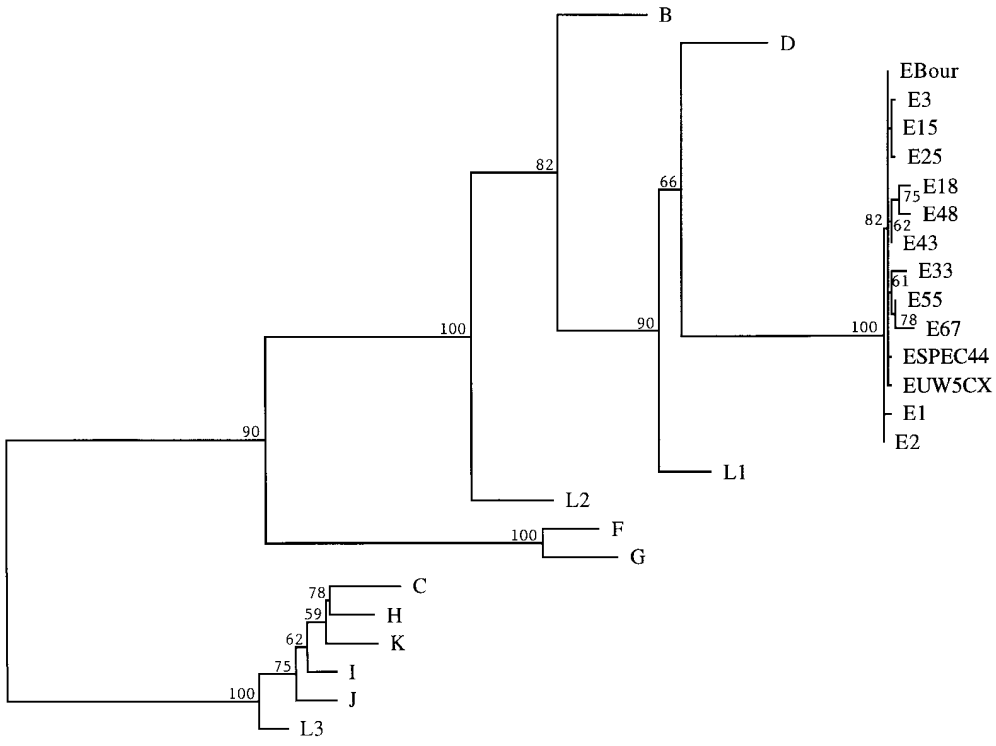

Figure 4. Phylogenetic analysis of representative E prototype genotypes, E variant genotypes, and genital prototype genotypes D, F, G, H, I, J, K, L1, L2, and L3. The relative branch lengths were drawn from the most parsimonious tree; branch point numbers represent the 50% threshold majority consensus values after 100 bootstrap replicates for weighted-parsimony. Note the two clades for E variants: E/18 and E/48, and E/55 and E/67.

Table II. Frequencies of E Signature-Pattern Amino Acids in Comparison with Prototype Genital Strains and E Variants

	64*								VS1											83	89	102	120	121
	-	D‡	K	P	.	.	T	T	G	N	A	T	A	P	T	.	L	T	-	.	.	E	R	C	S	S
E Sequence Set	1.0§	1.0	1.0	1.0			1.0	1.0	0.93	1.0	1.0	1.0	1.0	1.0	1.0		1.0	0.86	1.0			1.0	1.0	1.0	1.0	1.0
									E/67									E/55								
																		E/67								
Genital Sequence Set	0.0	0.0	0.22	0.22			0.11	0.22	0.22	0.22	0.11	0.0	0.22	0.22	0.0		0.44	0.22	0.22			0.22	0.22	0.44	0.11	0.44
			B‖	B			B	B	B	B	B		B	B			B	B	B			B	B	B	B	B
			D	D				D	D	D			D	D			D	D	D			D	D	D		D
																	F							F		F
																	G							G		G

	139								VS2											160	169	171	183	186	211
	D	N	E	N	Q	.	T	V	K	.	N	S	-	.	.	.	M	S	.	D	.	S	A	S	S	P	A
E Sequence Set	1.0	1.0	1.0	1.0	1.0		1.0	1.0	1.0		1.0	1.0	1.0				1.0	0.93		0.93		1.0	1.0	1.0	1.0	1.0	1.0
																		E/18		E/48							
Genital Sequence Set	0.33	0.22	0.33	0.44	0.22		0.11	0.22	0.11		0.0	0.11	0.11				0.22	0.22		0.33		0.44	0.11	0.0	0.0	0.0	0.11
	D	B	B	B	B		D	B	D			D	D				B	B		B		B	B				G
	F	D	D	D	D			D									D	D		D		D					
	G		G	F																K		F					
				G																		G					

	224		VS3		231	235	274	281	288		VS4	295	305				VS4			314	317	325
	Q	.	.	.	A	L	I	D	A	V	S	.	T	.	.	F	.	T	//	A	.	D	.	K	A	.	A	-	.	G	G	T
E Sequence Set	1.0			1.0	1.0	1.0	1.0	1.0	0.86	1.0	1.0		1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
									E/18																						
									E/48																						
Genital Sequence Set	0.11			0.11	0.44	0.11	0.44	0.44	0.0	0.22	0.22		0.22	0.11	0.22	0.22	0.22	0.33	0.33	0.33	0.44	0.22	0.0
	G			G	B	I	B	B		B	D		B	D	B	B	B	G	B	B	B	B	
					D		D	D		D	F		D		D	D	D	J	D	D	D	D	
					F		F	F										H	I	I	F		
					G		G	G													G		

*Numbers above line represent amino acid position on the major outer membrane protein. Variable segments are denoted as VS1, 2, 3, and 4. ‡Letters within boxes represent signature amino acids that define E as distinct from all other non-E genotypes. These were the most common amino acids for the 12 E genotypes in this paper; this is a majority signature. Underlined amino acids are in VS. Periods denote positions where amino acids are shared by all serovars within a VS. §Numbers with decimal points under boxes denote the frequency among all the 12 E genotypes for each amino acid in the signature and, where this differs from 1.0, the E sequence that is different from the other E genotypes is placed below the respective frequency. Below this are the frequencies for all other non-E genital genotypes in comparison with the E signature amino acids. Some of these match the amino acids of the E sequence set, but, for nine amino acids, none of the genital genotypes match, hence a 0.0 frequency is observed. ‖Capital letters denote serovars that share the amino acid of the E signature pattern.

product that was directly sequenced. The cloned sequences revealed a single molecular variant for each variant allele when compared with the direct PCR-based sequence. Thus, no variants were generated by either oligonucleotide primer pair.

Computational results. Nucleotide and amino acid alignments were generated for E prototype and variant genotypes and for prototype genital genotypes. From the phylogenetic analyses we found that four of the variants formed two distinct

subsets or clades within the group of E sequences: E/18 and E/48, and E/55 and E/67 (Fig. 4). Note the high bootstrap values for this analysis. We used VESPA to determine a nucleotide and amino acid signature-pattern representative of E sequences. We found a signature-pattern of 57 amino acids that differed between the most common amino acid found for the E sequence set and the most common amino acid observed for the non-E genital prototype sequence set which could reliably define an E strain. As part of this analysis, we were able to determine the frequency of concordance of the E variant sequences with the E sequence set. Four of the variants had two nonconservative amino acid substitutions each and thus did not entirely match all 57 amino acids of the E signature-pattern (Table II). These amino acids were located at positions 70 (E/67) and 79 (E/55, E/67) in VS1, 156 (E/18) and 158 (E/48) in VS2, and 281 (E/18, E/48) which is 7 amino acids upstream of VS4. In the latter position, no other genital strains matched the E genotype signature-pattern. This represents a frequency of 0.0 (Table II). Further, all of these four variants were the same variants that formed new clades on the phylogenetic tree.

Discussion

Allelic polymorphism of the single copy gene *omp1* is considered the genetic basis for antigenic change within *C. trachomatis* (12, 30). Considerable *omp1* variation has been described recently among trachoma and STD populations around the world (1, 2, 11–17). For example, 62–100% of I, J, and K genotypes have been found to be variants (1, 2, 12). This degree of change in *omp1* contradicts the previous notion that *omp1* sequences have been relatively stable for decades (31). Yet, it is important to consider that many of the newly identified genotypes may have been present for a long period of time in a population and only recently identified when the appropriate technology became available for genotyping. E genotypes have not displayed the same degree of *omp1* polymorphism. However, these data have been based on the partial *omp1* sequence of 57 E samples from different parts of the world (1, 2, 12, 13). In the current study, the complete *omp1* gene of 67 serovar E samples was sequenced and used for genotyping.

Only 11 (16%) distinct E variants were found among our study population. 8 (73%) of the 11 had nucleotide substitutions that resulted in 1–6 nonsynonymous codons, the majority of which were nonconservative amino acids (Figs. 1–3). Of the 39 nucleotide substitutions among the 11 variants, only 7 (18%) resulted in synonymous codons. Although the overall number of variants was low compared with those described for other serovars, an analysis of the nucleotide and amino acid substitutions suggests a similar trend in *omp1* diversification compared with other serovars. In the current data set, two to four variants were found to have the same amino acid change at the same position on MOMP compared with E/Bour (Figs. 1–3). This mutational pattern has also been observed for variants of other serovars (1, 2, 11–17). Interestingly, some of the mutations contained in the E variants resulted in the same amino acid found in closely related serovars: Leu 157 Phe for variants E/1, E/2, and E/48, and serovars D and L1 (Fig. 2; VS2), and Ala 291 Thr for E/25, E/43, sample 44, and serovars B, Ba, and L2 (Fig. 3; VS4). Further, there was a Leu 78 Arg for variants E/55 and E/67 (Fig. 1; VS1), yet, a Leu 78 Cys for serovars L1 and L2. Thus, certain mutations tend to occur at specific positions on MOMP which suggests evolutionary relatedness. Some investigators have observed that closely related serovars accumulate substitutions at specific locations and that more distantly related serovars accumulate deletions and insertions which might promote *omp1* diversification (12, 30). These positions, then, are thought to be "hot spots" for mutations which raises the question of whether these occur as random mutations or as a result of selection.

The high number of nucleotide substitutions that resulted in nonsynonymous codons (30/39) in this and other studies (1, 2, 11–17) suggests that selection may play a role in *omp1* diversification. Further, of the nonsynonymous codons in this study, 16 (53%) were nonconservative amino acids, most of which occurred in VS. Interestingly, we found no mutations in areas encoding recognized B cell epitopes (32). It is possible that our sample size was too small to detect these mutations. Yet, in another study (2), we sequenced 18 E genotypes (without paired immunotyping data) and these sequences contained a similar number and type of mutation as found in this study. In contrast, three of the variants (E/18, E/48, and E/67) contained one to three nonsynonymous codons in a region known to be a T cell epitope (Fig. 2). This epitope spans residues 191–251 which include VS3 (10). This is important as protection from reinfection with *C. trachomatis* is most likely determined by mucosal antibodies and is T cell dependent (33). Thus, mutations in this region may alter T cell determinants and the T cell response. A change in response may promote persistence of infection or facilitate disease progression which could significantly impact transmission.

Altering surface or T cell epitopes may not be the only mechanism used by this serovar to deal with host immune responses. For example, the avian influenza virus is quite unique in that it appears to have reached an equilibrium state with its host. The virus produces a large number of progeny but variants are uncommon and, over the last 60 yr, there has been no net evolutionary progress (34). The data we present here for serovar E would suggest that it may have evolved similarly to avian influenza virus. This proposition will have to be explored in in vitro and animal model studies comparing prototype and variant strains of serovar E with other serovars and by examining local immune responses.

We pursued a number of techniques to differentiate E serovars from all other serovars and to determine which E variants may represent biologically important strains versus those which are transiently or even persistently represented but of no lasting consequence for the organism. The results of immunotyping provided no additional information towards these goals. We then examined the *omp1* sequences of all E serovars from this and other studies and compared these with non-E prototype sequences. We found that the E sequences were identical in sequence to ocular (trachoma) but not genital Ba isolates in the constant region upstream of VS1. These findings were identical to those of Frost et al. (13). We would have expected the E genotypes to resemble the sequence of other urogenital genotypes, especially in constant regions. However, this similarity to trachoma strains upstream of VS1 may promote transmission of the organism by perhaps extending the tissue tropism of the organism.

Phylogenetic analysis of specific genes from a pathogen has historically been used to differentiate strain types especially when there is limited data regarding the functional differences among the strains (35). Given the limited functional data for variants of *C. trachomatis,* we conducted phylogenetic analyses

of *omp1* and MOMP as these appear to be sufficiently discriminating for typing *C. trachomatis* and, thus, for establishing phylogenies. Four variants (E/18 and E/48, and E/55 and E/67) formed two distinct clades within the E sequences on the tree (Fig. 4). Three of these four variants had six nonsynonymous codons, three of which were nonconservative amino acids. These variants contained the highest number of nonsynonymous codons for any variant and thus there was a higher expectation that they would form new clades. The data suggest that these four variants have evolved from E prototypes or other E variants and may represent distinct variants that are biologically significant.

In addition to the phylogenetic classification, we used signature-pattern analysis to evaluate amino acid sequence characteristics for differentiating E from all other non-E chlamydial strains and for distinguishing E variants. We found a signature-pattern of 57 amino acids for the E sequence set. This analysis considered all VS and conserved regions such that amino acid signature-patterns could be defined for the entire MOMP. Previously, prototype strains have been distinguished by differences in VS only. Thus, we were able to define with confidence a distinct signature-pattern that represents an E strain. Interestingly, 5 of the 10 amino acids with 0.0 frequencies (no amino acid matches with non-E strains) were outside of VS (Table II) suggesting that "constant regions" are just as important for defining an amino acid signature-pattern for E as a VS. We were also able to determine the frequency of concordance of the E variant sequences with the E sequence set. Four of the variants did not match the E signature-pattern by two nonconservative amino acid substitutions each (Table II). These four were the same variants that formed two new clades on the phylogenetic tree, and three of these four displayed nonsynonymous codons in an important T cell epitope. Since one of the major questions for chlamydia research is how to define a variant, these analyses offer for the first time a unique approach towards this goal. Thus, a new variant of E may be defined as one that has limited variation from the characteristic signature-pattern for E but is phylogenetically distinct.

Our study reveals that identical E immunotypes display sequence diversity within both constant and variable regions of *omp1*. These changes may contribute to protein conformation, epitope exposure, and altered infectivity, although additional studies will be required to address these issues. An operational definition of variants based on computational analyses of *omp1* and MOMP sequences will enhance our understanding of the evolution of *C. trachomatis.* This will aid in identifying distinct biological variants which will facilitate the investigation of functional differences relative to the type of amino acid substitutions found in the sequence. Further, an operational definition will be instrumental for clarifying the mutation trends, molecular epidemiology, and transmission dynamics of *C. trachomatis* which will be essential for designing appropriate public health interventions and an effective vaccine.

Acknowledgments

Thanks go to Maricelle Lancero and Thomas Bridges for excellent technical assistance, and Dr. Paul P. Ulrich for critical review of the manuscript.

This research was supported in part by The McCormick Fund, Stanford University School of Medicine, and Public Health Service grant EY00310 from the National Institutes of Health.

References

1. Yang, C.L., I. Maclean, and R.C. Brunham. 1993. DNA sequence polymorphism of the *Chlamydia trachomatis omp1* gene. *J. Infect. Dis.* 168:1225–1230.

2. Dean, D., E. Odens, G. Bolan, N. Padian, and J. Schachter. 1995. Major outer membrane protein variants of *Chlamydia trachomatis* are associated with severe upper genital tract infections and histopathology in San Francisco. *J. Infect. Dis.* 172:1013–1022.

3. Rodriguez, P., B. de Barbeyrac, K. Persson, B. Dutilh, and C. Bebear. 1993. Evaluation of molecular typing for epidemiological study of *Chlamydia trachomatis* genital infections. *J. Clin. Microbiol.* 31:2238–2240.

4. Kuo, C.-C., S.-P. Wang, K.K. Holmes, and J.T. Grayston. 1983. Immunotypes of *Chlamydia trachomatis* isolates in Seattle, Washington. *Infect. Immun.* 41:865–868.

5. Chungue, E., J. Soulier, G. Philippon, and S.-P. Wang. 1994. Immunotypes of *Chlamydia trachomatis* isolated from genital tract specimens in Tahiti. *Eur. J. Clin. Microbiol. Infect. Dis.* 13:436–438.

6. Centers for Disease Control and Prevention. 1993. Recommendations for the prevention and management of *Chlamydia trachomatis* infections. *Morb. Mortal. Wkly. Rep.* 42(RR-12):1–39.

7. Wang, S.-P., C.-C. Kuo, R.C. Barnes, R.S. Stephens, and J.T. Grayston. 1985. Immunotyping of *Chlamydia trachomatis* with monoclonal antibodies. *J. Infect. Dis.* 152:791–800.

8. Baehr, W., X.I. 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.

9. Stephens, R.S., E.A. Wagar, and G.K. Schoolnik. 1988. High-resolution mapping of serovar-specific and common antigenic determinants of the major outer membrane protein of *Chlamydia trachomatis*. *J. Exp. Med.* 167:817–831.

10. Allen, J.E., R.M. Locksley, and R.S. Stephens. 1991. A single peptide from the major outer membrane protein of *Chlamydia trachomatis* elicits T cell help for the production of antibodies to protective determinants. *J. Immunol.* 147:674–679.

11. Lampe, M.F., R.J. Suchland, and W.E. Stamm. 1993. Nucleotide sequence of the variable domains within the major outer membrane protein gene from serovariants of *Chlamydia trachomatis*. *Infect. Immun.* 61:213–219.

12. Brunham, R., C. Yang, I. Maclean, J. Kimani, G. Maitha, and F. Plummer. 1994. *Chlamydia trachomatis* from individuals in a sexually transmitted disease core group exhibit frequent sequence variation in the major outer membrane protein (*omp1*) gene. *J. Clin. Invest.* 94:458–463.

13. Frost, E.H., S. Deslandes, D. Gendron, D. Bourgaux-Ramoisy, and P. Bourgaux. 1995. Variation outside variable segments of the major outer membrane protein distinguishes trachoma from urogenital isolates of the same serovar of *Chlamydia trachomatis*. *Genitourin. Med.* 71:18–23.

14. Dean, D., J. Schachter, C.R. Dawson, and R.S. Stephens. 1992. Comparison of the major outer membrane protein variant sequence regions of B/Ba isolates: a molecular epidemiologic approach to *Chlamydia trachomatis* infections. *J. Infect. Dis.* 166:383–392.

15. Dean, D., M. Patton, and R.S. Stephens. 1991. Direct sequence evaluation of the major outer membrane protein gene variant regions of *Chlamydia trachomatis* subtypes D′, I′, L2′. *Infect. Immun.* 59:1579–1582.

16. Hayes, L.J., R.L. Bailey, D.C.W. Mabey, I.N. Clarke, M.A. Pickett, P.J. Watt, and M.E. Ward. 1992. Genotyping of *Chlamydia trachomatis* from a trachoma-endemic village in The Gambia by a nested polymerase chain reaction: identification of strain variants. *J. Infect. Dis.* 166:1173–1177.

17. Hayes, L.J., S. Pecharatana, R.L. Bailey, T.J. Hampton, M.A. Pickett, D.C.W. Mabey, P.J. Watt, and M.E. Ward. 1995. Extent and kinetics of genetic change in the *omp1* gene of *Chlamydia trachomatis* in two villages with endemic trachoma. *J. Infect. Dis.* 172:268–272.

18. Dean, D. 1994. Molecular characterization of new *Chlamydia trachomatis* serological variants from a trachoma endemic region of Africa. *In* Chlamydial Infections. J. Orfila, G.I. Byrne, M.A. Chernesky, J.T. Grayston, R.B. Jones, G.L. Ridgway, R. Saikku, J. Schachter, W.E. Stamm, and R.S. Stephens, editors. Societa Editrice Esculapio, Bologna. 259–262.

19. Wang, S.P., and J.T. Grayston. 1991. Three new serovars of *Chlamydia trachomatis*: Da, Ia, and L2a. *J. Infect. Dis.* 163:403–405.

20. Stephens, R.S. 1993. Challenge of *Chlamydia* research. *Infect. Agent. Dis.* 1:279–293.

21. Ripa, K.T., and P.-A. Mardh. 1977. Cultivation of *Chlamydia trachomatis* in cyclohexamide treated McCoy cells. *J. Clin. Microbiol.* 6:328–331.

22. Dean, D., and R.S. Stephens. 1994. Identification of individual genotypes of *Chlamydia trachomatis* from experimentally mixed serovars and mixed infections among trachoma patients. *J. Clin. Microbiol.* 32:1506–1510.

23. Smith, R.F., and T.F. Smith. 1990. Automatic generation of primary sequence patterns from sets of related protein sequences. *Proc. Natl. Acad. Sci. USA.* 87:118–122.

24. Faulkner, D.V., and J. Jurka. 1988. Multiple aligned sequence editor (MASE). *Trends Biochem. Sci.* 13:321–322.

25. Maddison, W.P., and D.R. Maddison. 1995. MacClade: Analysis of Phylogenetic and Character Evolution. Sinauer Associates, Inc., Sunderland, MA.

26. Felsenstein, J. 1989. PHYLIP-phylogeny inference package (version 3.2). *Cladistics.* 5:164–166.

27. Korber, B.T.M., and G. Myers. 1992. Signature pattern analysis: a method for assessing viral sequence relatedness. *AIDS Res. Hum. Retro.* 8: 1549–1560.

28. Korber, B.T.M., K.J. Kunstman, B.K. Patterson, M. Furtado, M.M. McEvilly, R. Levy, and S.M. Wolinsky. 1994. Genetic differences between blood- and brain-derived viral sequences from human immunodeficiency virus type 1-infected patients: evidence of conserved elements in the V3 region of the envelope protein of brain-derived sequences. *J. Virol.* 68:7467–7481.

29. Eckert, K.A., and T.A. Kunkel. 1990. High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Nucleic Acids Res.* 18:3739–3744.

30. Stephens, R.S., R. Sanchez-Pescador, E.A. Wagar, C. Inouye, and M.S. Urdea. 1987. Diversity of *Chlamydia trachomatis* major outer membrane protein genes. *J. Bacteriol.* 169:3879–3885.

31. Morrison, R.P., D.S. Manning, and H.D. Caldwell. 1992. Immunology of *Chlamydia trachomatis* infections: immunoprotective and immunopathogenic responses. *In* Sexually Transmitted Diseases. Vol. 8. T.C. Quinn, editor. Raven Press, New York. 57–84.

32. Batteiger, B.E. 1996. The major outer membrane protein of a single *Chlamydia trachomatis* serovar can possess more than one serovar-specific epitope. *Infect. Immun.* 64:542–547.

33. Rank, R.G. 1988. Role of the immune response. *In* Microbiology of Chlamydia. A.L. Barron, editor. CRC Press Inc., Boca Raton, FL. 209.

34. Webster, R.G., W.J. Bean, O.T. Gorman, T.M. Chambers, and Y. Kawaoka. 1992. Evolution and ecology of influenza A viruses. *Microbiol. Rev.* 56:152–179.

35. Chan, S.-Y., H. Delius, A.L. Halpern, and H.-U. Bernard. 1995. Analysis of genomic sequences of 95 papilloma virus types: uniting typing, phylogeny, and taxonomy. *J. Virol.* 69:3074–3083.

36. 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.* 577:1040–1049.