INFECTION AND IMMUNITY, Aug. 1998, p. 3618–3625 Vol. 66, No. 8
0019-9567/98/$04.00+0

Phylogenetic Analysis of the *Chlamydia trachomatis* Major Outer Membrane Protein and Examination of Potential Pathogenic Determinants

DIANE R. STOTHARD,* GEORGE BOGUSLAWSKI,† AND ROBERT B. JONES

Department of Medicine, Indiana University School of Medicine, Indianapolis, Indiana 46202

Received 21 November 1997/Returned for modification 24 March 1998/Accepted 26 May 1998

Phylogenetic analysis was utilized to investigate biological relationships (tissue tropism, disease presentation, and epidemiologic success), as evidenced by coevolution, among human strains of *Chlamydia trachomatis*. Nucleotide sequences of *omp1*, the gene encoding the major outer membrane protein (MOMP) of *C. trachomatis*, were determined for 40 strains representing 11 serovars. These data were combined with available *omp1* sequences from GenBank for an analysis encompassing a total of 69 strains representing 17 serovars infecting humans. Phylogenetic analysis of the nucleotide and inferred amino acid sequences showed no evolutionary relationships among serovars that corresponded to biological or pathological phenotypes (tissue tropism, disease presentation, and epidemiologic success). In addition, no specific residues that may have evolved to play a role in determining biologically relevant characteristics of chlamydia, such as tissue specificity, disease presentation, and epidemiologic success, were apparent in the MOMP. These results suggest that variation in MOMP may have arisen from a need to be diverse in the presence of immune pressure rather than as a function of pathogenicity. Therefore, the role of MOMP in disease pathogenesis and infection may be passive, and it may not be the major ligand responsible for directing infection of various human cell types.

Chlamydia trachomatis causes many human diseases, including ocular trachoma (serotypes A to C); urethritis, epididymitis, cervicitis, and salpingitis (serotypes D to K); and lymphogranuloma venereum (LGV; serotypes L1 to L3) (41–43). Serotypes A to K (trachoma biovar) primarily infect columnar epithelial cells of the mucous membranes, while serotypes L1 to L3 (LGV biovar) also proliferate in lymphatic tissue and cause a more systemic infection. The third biovar of *C. trachomatis* is the murine biovar, consisting of one strain, MoPn, which causes mouse pneumonitis but does not infect humans. In addition, three serogroups among the human biovars that appear to be independent of biovar have been defined: the B complex (serotypes B, Ba, D, Da, E, L1, L2, and L2a), the C complex (serotypes A, C, H, I, Ia, J, K, and L3), and the intermediate group (serotypes F and G) (9, 42, 44).

Serotype specificity is conferred by the major outer membrane protein (MOMP; the product of the *omp1* gene). MOMP constitutes ~60% of the protein mass of the chlamydial outer membrane and has been shown to have porin-like characteristics in vitro (5). MOMP is thought to play a role in the structural integrity of the organism (7, 8, 19) and is surface exposed and glycosylated (2, 10, 25, 39). MOMP contains 7 to 10 cysteines which may form homo- or hetero-oligomers with itself and/or other outer membrane proteins (28, 30).

The amino acid sequence of MOMP exhibits heterogeneity that is mainly localized to four hypervariable segments (VS1 to VS4) (44) which are surface exposed and reactive with human immune sera (45, 46). Monoclonal antibodies directed against MOMP are neutralizing in cell culture and in some animal models (2, 27, 31, 45, 46). Although limited, protective immunity is serovar specific, making MOMP a focus of vaccine development.

The outer membrane proteins of obligate intracellular bacteria play a direct role in the process of adaptation by facilitating interactions between the bacterial cell and its host cell. The surface of the chlamydial elementary body must provide components responsible for (i) protection against the environment outside of the host, (ii) defense against host immune response, (iii) attachment to host cells, and (iv) prevention of phagosome-lysosome fusion. Variability in MOMP sequence is presumably a result of host selection and bacterial adaptation. Thus, MOMP has been implicated in the mechanisms of attachment, infection, and/or pathogenesis because of its variability and its exposed location. Consequently, an evolutionary examination of the *omp1* gene and the MOMP may provide insight into the role of MOMP in the processes of infection.

Fitch et al. (16) examined the *omp1* gene by sequencing VS regions of 15 serotypes (17 strains) and the complete sequence of 9 serotypes (11 strains). The authors concluded that there was no evidence for coevolution of human strains of *C. trachomatis* and the human host. However, they did not examine MOMP at the amino acid level, and their study was limited to one or two strains from each serovar. The present study combines our *omp1* sequences (40 strains from 11 serovars) with *omp1* sequences available in GenBank for an analysis encompassing 69 strains from 17 serovars. Our goal was to examine the patterns of substitution in MOMP at both the nucleic acid level (*omp1*) and the amino acid level (MOMP) to identify residues that may be correlated with serovar-specific tissue tropism or pathogenic potential by virtue of shared evolutionary divergence.

MATERIALS AND METHODS

Strains. Sequences of the MOMPs derived from strains used in this study were analyzed along with sequences from strains available in GenBank (Table 1). Several were epidemiologically linked by being isolates from individuals in a sexual partnership or isolates from the same individual at different times.

* Corresponding author. Mailing address: Department of Medicine, Infectious Diseases, Indiana University School of Medicine, 435 Emerson Hall, 545 Barnhill Dr., Indianapolis, IN 46202. Phone: (317) 278-1791. Fax: (317) 274-1587. E-mail: dstothar@iupui.edu.

† Present address: Department of Biochemistry, Indiana University School of Medicine, Indianapolis, IN 46202.

TABLE 1. Strains

Serovar/strain	Yr of isolation	Source	Reference and/or comments	GenBank accession no.
A/Sa-1	1957	Conjunctiva	20	M58938
A/Har-13	1958	Conjunctiva	2	J03813
B/TW-5	1959	Conjunctiva	38	M17342
B/Jali-20	1985	Conjunctiva	21	M33636
Ba/Apache-2[a]	1960	Conjunctiva	This study	AF063194
B/Alpha-95	1996?	Urethra	13	U80075
B/IU-1226	1987	Urine	This study	AF063208
C/TW3	1959	Conjunctiva	38	M17343
D/B120	1983–1991	STD[e]	35	X62918
D/B185	1983–1991	STD	35	X62919
D/IU-8[b]	1989	Urethra	This study; identical to D/B120	See X62918
D/IU-19[b]	1989	Cervix	This study; identical to D/B120	See X62918
D/IU-71960	1987	Urethra	This study	AF063195
D/IU-72403	1987	Cervix	This study	AF063196
D/IU-83786	1988	Urethra	This study	AF063197
D/IU-80021	1987	Cervix	Identical to D/IU-83786	AF063197
D/IU-88712	1989	Cervix	Identical to D/IU-71960	AF063195
D/IU-38638	1985	Cervix	Identical to D/IU-71960	AF063195
D/IU-87178	1989	Cervix	Identical to D/IU-71960	AF063195
D/IU-83836	1988	Cervix	Identical to D/IU-83786	AF063197
D/IU-47674	1985	Cervix	Identical to D/IU-83786	AF063197
D/IU-86390	1988	Urethra	Identical to D/IU-83786	AF063197
D/IC-Cal8	1991?	STD	35	X62920
Da/TW-448	1991?	STD	35	X62921
Da/EV-293	1983–1991	STD	(36)	X77365
Da/IU-1554	1997	Urine	This study; identical to D/IC-Cal8	AF063209
E/UW5	?	Cervix	This study; identical to E/Bour (32)	See X52557
E/Bour 1990	1959	Conjunctiva	32	X52557
E/Bour 1997	1959	Conjunctiva	12	U78763
E/1	1989–1992	STD	12	U78528
E/2	1989–1992	STD	12	U78529
E/3	1989–1992	STD	12	U78530
E/15	1989–1992	STD	12	U78531
E/18	1989–1992	STD	12	U78532
E/25	1989–1992	STD	12	U78533
E/33	1989–1992	STD	12	U78534
E/43	1989–1992	STD	12	U78535
E/48	1989–1992	STD	12	U78536
E/55	1989–1992	STD	12	U78537
E/67	1989–1992	STD	12	U78538
E/IU-1579	1997	Urine	This study; identical to E/Bour (32)	AF063211
E/IU-1586	1997	Urine	This study; identical to E/Bour (32)	AF063210
E/IU-51538[c]	1986	Cervix	This study	AF063198
E/IU-71253[c]	1987	Urethra	Identical to E/IU-51538	AF063198
E/IU-71254[c]	1987	Cervix	Identical to E/IU-51538	AF063198
F/IC-Cal3	1960?	Conjunctiva	47	X52080
F/IU-1552	1997	Urine	This study	AF063212
F/IU-1607	1997	Urine	This study; identical to F/IC-Cal3	AF063213
G/UW57[a]	1971	Cervix	This study	AF063199
H/UW4	1965	Cervix	18	X16007
I/UW12[a]	1966	Urethra	This study	AF063200
Ia/IU-4168[a]	1987	Urethra	This study	AF063201
Ia/IU-1588	1997	Urine	This study	AF063205
J/UW36[a]	1971	Cervix	This study	AF063202
J/IU-1553	1997	Urine	This study; identical to J/UW36	AF063206
Ja/IU-A795	1986	Cervix	This study	AF063203
Ja/IU-A1300	1986	Cervix	Identical to Ja/IU-A795	AF063203
Ja/IU-60063	1986	Urethra	Identical to Ja/IU-A795	AF063203
Ja/IU-71171	1987	Urethra	Identical to Ja/IU-A795	AF063203
Ja/IU-37538	1985	Cervix	Identical to Ja/IU-A795	AF063203
Ja/IU-51960	1986	Cervix	Identical to Ja/IU-A795	AF063203
Ja/IU-115718	1992	Cervix	Identical to Ja/IU-A795	AF063203
Ja/IU-142018[d]	1994	Cervix	Identical to Ja/IU-A795	AF063203
Ja/IU-142019[d]	1994	Urethra	Identical to Ja/IU-A795	AF063203
K/UW31	1973?	Cervix	This study	AF056204
K/IU-1619	1997	Urine	This study; identical to K/UW31	AF063207
L1/440	1968	Bubo	33	M36533
L2/434	1968	Bubo	37	M14738
L3/404	1967	Bubo	15	X55700
MoPn/Nigg II	1939	Lung (mouse)	15	M64171

[a] Partial sequence for strain described is available in GenBank.
[b] These strains have identical sequences; they were taken from a sexual contact pair.
[c] Strains E/IU-51538 and E/IU-71254 are from the same patient at different time points, while E/IU-71253 is from the sexual contact of this patient at the second time point.
[d] These strains are from the same patient but different body sites.
[e] STD, sexually transmitted disease.

Cell culture. Clinical swab specimens were collected and expanded in cell culture as previously described (24). To identify inclusions, immunofluorescence staining was done by using a genus-reactive monoclonal antibody (23). Expansion of each isolate was done via serial passage on glass coverslips in shell vials by using antibiotic-free medium (24).

DNA isolation from cultured cells. All tips, tubes, and buffers were UV irradiated to reduce contaminating DNA. Infected McCoy cell monolayers were disrupted and transferred to sterile 2-ml tubes. One hundred to 200 μl of the cell suspension was centrifuged for 5 min, the supernatant was removed, and the cell pellet was resuspended in 500 μl of sodium dodecyl sulfate (SDS) lysis buffer (0.5 M Tris-EDTA [pH 7.5], 1% SDS, 0.1 mg of proteinase K per ml). The suspension was incubated at 55°C for 2 h and at 95°C for 10 min. The resulting lysate was extracted with an equal volume of phenol-chloroform. DNA was precipitated in an equal volume of isopropanol, washed one time with 70% ethanol, dried, and resuspended in 50 μl of UV-irradiated, double-distilled water.

Preparation of urine specimens. First-void urine was collected in sterile cups and stored immediately at 4°C until processed. Precipitates were resuspended by warming the urine at 37°C for 15 to 30 min before processing. Two milliliters of urine was microcentrifuged at high speed (12,000 × *g*) for 10 min. The supernatant was discarded, and 1 ml of transport medium was used to wash the pellet. The cells were centrifuged again, for 5 min at high speed (12,000 × *g*). The supernatant was discarded, and the pellet was resuspended in 200 μl of SDS lysis buffer. DNA extraction and isolation were carried out as described above.

DNA amplification. Primers specific to the *C. trachomatis* MOMP were based on published sequences (Table 1). The 5′ primer, MOMP-108, (5′-GGC CAT TAA TTG CTA CAG GAC ATC TTG TC-3′) is located 108 bp upstream of the serovar A *omp1* gene in the 5′ noncoding region. The 3′ primer, RVS1163, (5′-CGG AAT TGT GCA TTT ACG TGA G-3′) is located at bp 1163 in the serovar A *omp1* gene (11). For samples which required reamplification, the following nested primers were designed: MOMP87 (5′-TGA ACC AAG CCT TAT GAT CGA CGG A-3′) and RVS1059 (5′-GCA ATA CCG CAA GAT TTT CTA GAT TTC ATC-3′).

All tips, tubes, and buffers were UV irradiated to reduce contaminating DNA. One negative control reaction mixture was run for every seven experimental reaction mixtures. DNA amplification was done by using the PCR Core Kit plus uracil-DNA-glycosylase (UNG) (Boehringer Mannheim, Indianapolis, Ind.). One-hundred-microliter-volume reaction mixtures consisted of 5 to 10 μl of template DNA, 30 pmol of each primer, 2 mM $MgCl_2$, 200 μM each deoxynucleoside triphosphate and 600 μM dUTP, 2 U of heat-labile UNG, and 2.5 U of *Taq* polymerase in a 1× reaction buffer. UNG PCR mixtures were incubated at room temperature for 10 min for the UNG reaction to proceed. Cycling conditions began with an initial 5-min denaturation step at 95°C which also inactivated the UNG, followed by 40 cycles of denaturation (95°C, 30 s), annealing (60°C, 30 s), and extension (72°C, 1 min). An additional 10-min extension at 72°C was performed at the end of the 40 cycles. Once verified by agarose gel electrophoresis and ethidium bromide staining, replicate reaction mixtures were pooled for sequence analysis.

Nested reamplifications were done when amplifications did not produce sufficient yield for sequencing. The first round of amplification was done with 100-μl volumes as described above, except only for 30 cycles of amplification; this was followed by a nested reamplification of the first-round PCR mixture. Negative controls from the first round of amplification were also used as a template in the nested-reamplification negative controls. The PCR master mix was made as described above, and 95 μl was aliquoted into reaction tubes. The tubes were incubated at room temperature for 10 min and then at 95°C for 2 min to inactivate the UNG, after which 5 μl from the first-round PCR mixture was added to the tubes while they were maintained at 95°C. The reamplification consisted of the same cycling conditions as those of the first round, including the initial 5-min denaturation at 95°C, but was carried out for 40 cycles.

DNA sequencing. Direct sequencing of PCR products was done by using the BRL dsCycle sequencing kit (Life Technologies, Gaithersburg, Md.) with Isoblue [γ-^{32}P]ATP (ICN, Costa Mesa, Calif.)-end-labeled sequencing primers. Reaction products were separated on a 6 to 8% polyacrylamide gel and visualized by autoradiography (Kodak BioMax XR). PCR products were also cloned with the TA cloning kit (pCR 2.1; Invitrogen, La Jolla, Calif.). To ensure sequence fidelity and to resolve ambiguities, sequence was obtained from at least two individual clones and compared to sequence obtained directly from PCR products. Therefore, each gene was sequenced in part at least three times. Several PCR products were also sequenced by using the ABI Prism 373 (Applied Biosystems) automated sequencing system employing dye terminators. Sequence obtained in this way was also collected from both strands to ensure sequence fidelity.

Nucleotide and amino acid sequence alignment. Both nucleotide and amino acid sequences were aligned with the assistance of the program ESEE (6). For regions of high variability, the alignment was based primarily on the most conservative amino acid substitutions and secondarily on nucleotide substitutions. The entire *omp1* sequence was not obtained for nine isolates for which nested reamplifications were done because of the intragene location of the nested primer pair. These sequences and partial sequences from GenBank were compared to other sequences but were not included in the phylogenetic analysis. Consequently, the full sequence alignment contained 29 strains and, for each strain, 1,165 characters (nucleotides plus alignment gaps) and 388 characters (amino acids plus alignment gaps). The alignment consisted of the entire *omp1* gene sequence except for the last 26 bases (and corresponding eight amino acids) which overlap the RVS1163 primer sequence and the last four nucleotides of the gene, not including the stop codon. This region is known for some of the serovars but not all and was excluded from the phylogenetic analysis.

Nucleotide and amino acid comparisons. After sequences were aligned, full-length nucleotide and amino acid sequences were compared manually in a pairwise fashion for differences. Contiguous insertions or deletions (IN/DEL) at the nucleotide and amino acid levels were counted as individual differences, i.e., a deletion of six nucleotides between two pairs was counted as six differences rather than one deletion event, and the resulting two-amino-acid deletion was counted as two deletions rather than as one deletion event. Pairwise comparisons were also made for strains with partial sequences, and differences were noted but not tallied.

Phylogenetic analysis. Residues corresponding to the RVS1163 primer and all regions of ambiguous alignment (gap IN/DEL positions) were excluded from phylogenetic analysis. Only the coding region was considered. Consequently, 1,128 of 1,165 nucleotides and 376 of 388 amino acids were compared. All phylogenetic reconstructions were carried out by using the Phylip package version 3.5c (14). Once the sequences were aligned, the corrected nucleotide divergence of them was calculated by using DNADIST, incorporating the Kimura two-parameter correction for multiple substitutions at a single site. The branching pattern of taxa was estimated from the distance matrix by neighbor-joining (34) by using the program NEIGHBOR. A cladistic reconstruction of the phylogenetic relationships among taxa was also performed by using the parsimony program DNAPARS. Phylogenetic reconstructions based on the inferred amino acid sequence of MOMP were also performed, by using PROTDIST, with the Kimura formula, and PROTPARS, with threshold parsimony. Confidence levels for the branching pattern were estimated by a bootstrap resampling of the data. Bootstrap values for the trees were obtained from a consensus tree based on 1,000 randomly generated data sets by using SEQBOOT with jumbled sequence addition. Analyses were performed on an IRIX Indigo (Silicon Graphics) UNIX workstation. The mouse pneumonitis strain, MoPn, was used as the outgroup in all analyses to place the root joining the human strains of *C. trachomatis*.

RESULTS

Sequence comparisons. The nucleotide and amino acid sequence comparisons for full-length sequences are summarized in Table 2. All sequence alignments are available upon request (to D.R.S.). Several strains had identical sequences (Table 1).

Six clinical strains that typed with our serovar J-specific monoclonal antibody (29) in microimmunofluorescence analysis had *omp1* sequences that were identical to each other but different from those of the prototype J strain (J/UW36) by 16 nucleotides and eight amino acids. Since these strains did react with the J antibody but were variant in their *omp1* sequence, they have been labeled Ja strains. Genotype analysis of nine other J-typed strains from the Indianapolis, Ind., area showed that three were of this genotype as well, for a total of nine isolates with the variant genotype (Table 1). Therefore, this genotype may be common in the local Indianapolis population and elsewhere. Evaluation of patient history information showed that the individuals from whom these isolates came were not known to be sexual contacts of one another.

Eleven of 13 serovar D isolates from the local Indianapolis population also differed from the prototype D strain (strain D/B120) (Table 1). Four isolates had only a single synonymous mutation compared to the prototype and thus were not included in the phylogenetic analysis: D/IU-71960, -88712, -83638, and -87178. One isolate, D/IU-1554, was identical to D/IC-Cal8, which was described by Sayada et al. (35). However, these two strains differed from the prototype serovar D strain by 10 nucleotides and four amino acids. The other 10 serovar D strains were not identical to previously published D or Da strains.

One isolate from urine, B/IU-1226, had a nucleotide sequence similar to that of the Bb genotype reported by Frost et al. (17). The sequence of B/IU-1226 differed from that of the reported Ba/Apache-2 and B/Jali20 ocular strains by 11 nucleotides and four amino acid substitutions. Isolate 1226 bound a serovar B- and Ba-specific monoclonal antibody but not any other serotype-specific antibody.

TABLE 2. Differences in the *omp1* and MOMP sequences between strains of *C. trachomatis*[a]

Serovar/strain	Nucleotide or amino acid difference with strain:																										
	B/TW-5	B/Jali20	Ba/Apache-2	B/Alpha-95	L2/434	D/IU-72403	D/IU-83786	D/B185	D/B120	Da/EV293	Da/TW448	D/IC-Cal8	L1/440	E/Bour 1990	F/IC-Cal3	G/UW57	A/SA1	A/Har-13	C/TW3	H/UW4	I/UW12	Ia/IU-A4168	J/UW36	Ja/IU-A795	K/UW53	L3/404	MoPn
B/TW-5		13	13	12	69	74	74	76	75	76	71	70	74	93	181	193	212	212	219	219	217	214	214	215	219	209	266
B/Jali20	6		6	7	70	66	66	68	67	68	64	63	69	91	178	190	209	209	215	215	214	211	211	212	215	206	259
Ba/Apache-2	5	6		7	70	69	69	71	70	71	66	65	67	89	176	188	219	219	217	217	214	212	212	211	217	207	256
B/Alpha-95	5	6	6		71	66	66	68	67	68	63	62	70	92	175	189	220	218	218	218	215	213	213	214	218	208	257
L2/434	24	25	25	24		93	92	95	94	95	89	88	73	93	171	162	212	212	222	221	220	218	217	218	223	208	256
D/IU-72403	24	23	25	24	29		2	5	4	5	11	8	44	77	168	181	218	218	212	208	214	212	217	218	212	212	255
D/IU-83786[b]	24	23	25	24	28	2		5	4	5	9	6	43	77	168	181	218	218	212	208	214	212	217	218	212	212	255
D/B185	23	22	24	23	28	3	3		1	2	12	11	46	79	167	180	207	207	209	205	212	209	214	215	209	212	255
D/B120	22	21	23	22	27	2	2	1		1	11	10	45	78	167	180	208	208	210	206	213	210	215	216	210	213	255
Da/EV293	23	22	24	23	28	3	3	2	1		10	11	46	79	166	180	209	209	209	205	213	209	214	215	209	212	255
Da/TW448	23	23	24	23	27	5	3	6	5	4		3	40	74	165	183	207	207	215	213	213	209	210	211	215	209	250
D/IC-Cal8	22	22	23	22	26	4	2	5	4	5	1		39	73	164	181	206	206	216	214	213	210	211	212	216	210	251
L1/440	26	26	25	27	23	18	17	17	16	17	16	15		75	155	176	205	205	219	222	218	216	212	213	220	208	256
E/Bour 1990[b]	23	24	22	23	28	22	22	21	20	21	21	20	20		179	166	215	216	226	230	225	224	220	226	228	220	256
F/IC-Cal3	60	61	59	58	58	55	55	54	54	53	52	53	53	57		39	228	230	234	235	235	230	226	229	240	231	274
G/UW57	57	57	55	55	56	55	55	54	54	54	53	53	54	51	16		235	237	233	234	235	231	232	234	238	236	280
A/SA-1	57	58	61	60	56	59	59	55	55	56	56	55	54	58	64	62		5	59	53	41	45	39	45	52	49	280
A/Har-13	59	60	63	60	58	61	61	57	57	58	58	57	56	60	66	64	4		62	54	43	47	40	46	53	50	281
C/TW3	67	67	68	67	66	65	65	63	62	62	64	65	63	67	70	69	29	30		34	37	31	32	29	35	45	285
H/UW4	60	60	61	60	60	56	56	54	54	54	58	58	58	64	69	66	27	29	22		29	20	23	33	24	35	280
I/UW12	57	58	58	57	56	58	58	56	56	56	56	56	55	58	66	65	19	21	20	17		13	22	28	28	33	274
Ia/IU-A4168	58	59	59	58	58	58	58	56	56	56	56	56	56	61	65	62	22	24	19	12	8		14	23	25	28	276
J/UW36	57	58	58	57	58	57	57	55	55	55	55	55	54	58	64	61	20	21	18	14	12	9		15	24	26	278
Ja/IU-A795[b]	60	61	60	60	60	60	60	58	58	58	58	58	56	61	66	64	23	24	18	17	13	11	8		34	30	279
K/UW53	59	59	60	59	61	58	57	55	55	55	57	57	58	62	71	67	26	27	20	18	18	18	14	18		38	285
L3/404	58	58	59	58	57	58	57	55	55	55	57	57	57	62	65	62	22	23	20	17	14	13	13	14	18		283
MoPn	73	73	72	72	70	72	72	71	71	71	71	71	71	69	76	72	74	76	76	75	72	71	72	74	75	76	

[a] Full-length sequences only. Numbers above the diagonal are nucleotide differences. Numbers below the diagonal are amino acid differences.
[b] D/IU-83786, E/Bour 1990, and Ja/IU-A795 are representatives of multiple strains with identical sequences. See Table 1 and Results.

The complete *omp1* sequence of E/Bour has been reported twice (12, 32). The two sequences differ by 10 bp and four amino acids even though they are purported to be two stocks of the same isolate. The sequences of our own stocks of E/Bour and E/UW5 were identical to the sequence published by Peterson et al. (32), and only 3 of 19 serovar E isolates sequenced for this study had an *omp1* sequence that differed by only one base from the original published E/Bour sequence (32).

Of 1,197 total sites in the human *omp1* alleles, ~790 are invariant. The LGV strains representing the three known serovars had a unique nucleotide at only two sites in the gene. At base 30, the LGV strains had a G while all other strains had an A. At base 117, the LGV strains had an A while all other strains had a G. Both of these substitutions were in conserved segment 1 and were silent. There were no other nucleotide substitutions or corresponding unique amino acid residues that were characteristic for ocular trachoma, LGV, or urogenital strains.

An examination of the amino acid alignment revealed eight highly conserved cysteine residues. Counting the initial methionine as amino acid 1, cysteine residues were at positions 48, 51, 55, 139, 207, 209, 232, and 360. All were located in the conserved segments of the corresponding gene sequence in all serovars, including the MoPn serovar. There were, however, some exceptions to this finding. Strain B/TW-5 (38) had a GTT codon (valine) instead of a cysteine TGT at residue 55, and the E/Bour strain reported in 1997 (E/Bour 1997) (12) along with all the serotype E variants described by Dean and Millman (12) had a TCC codon (serine) instead of a cysteine TGC codon corresponding to residue 207. There also were unique cysteine residues in some serovars in addition to the eight conserved cysteines. Serovars L1 and L2 had a cysteine at residue 103; B-complex strains (except serovar L2), and serovars F and G had a cysteine at residue 126; and serovars F and G had a cysteine at 329.

Phylogenetic analysis of the *omp1* gene. Figure 1 shows the distance neighbor-joining tree based on the nucleotide sequences of the *omp1* gene with consensus bootstrap values indicated at the nodes to which they apply. Bootstrap values lower than 50 are considered insignificant and are not shown. The most ambiguous regions of the tree are also those regions which join closely related taxa, namely, C-complex strains (Fig. 1). Both parsimony and neighbor-joining analyses produced similar trees, with some discrepancy in branching order of C-complex strains. Both analyses produced trees which segregated the serovars of *C. trachomatis* into three clades consistent with serocomplexes. Parsimony analysis of the *omp1* produced eight equally parsimonious trees (data not shown). Five of these trees differed in the branching order of D isolates. The other two trees differed in the branching order of strains within the C complex.

Phylogenetic analysis of the MOMP. Figure 2 presents the consensus neighbor-joining tree based on the amino acid sequences of the MOMP with bootstrap values at the nodes to which they apply. Parsimony analysis of the protein data produced 42 equally parsimonious trees, all of which indicated an early divergence of serovars F and G, prior to the divergence of the B- and C-complex strains. This branching pattern also agreed with the consensus neighbor-joining tree. However, the neighbor-joining analysis of the MOMP data set (Fig. 2) produced a tree where the B- and C-complex strains diverged first, prior to the divergence of the B-complex strains, similar to the nucleotide tree.

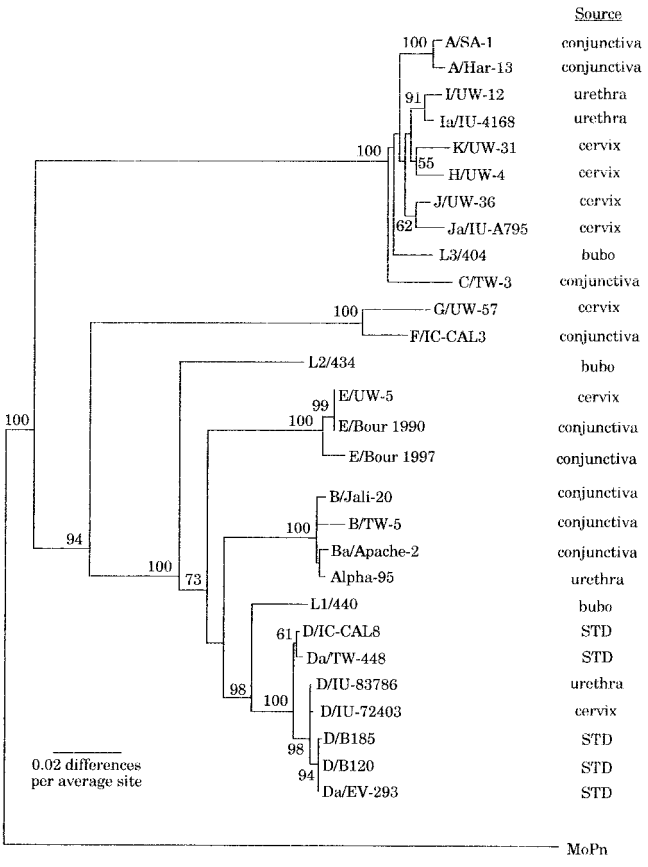

FIG. 1. The distance neighbor-joining tree based on the *C. trachomatis omp1* nucleotide sequence from 29 strains. The tree was rooted by using the *C. trachomatis* MoPn strain. Branch lengths are proportional to the amount of sequence divergence between taxa in the tree, as illustrated by the bar. Bootstrap values, based on 1,000 bootstrap replicates, are given at the nodes to which they apply. Bootstrap values below 50 are considered insignificant and are not shown.

DISCUSSION

Lampe et al. (26) recently showed that a MOMP variant could escape neutralization by both serovar-specific monoclonal antibodies and human immune sera. If this adaptive process occurs, the patterns of substitution in MOMP among serovars should be reflective of the biological relationships each serovar has with host cells. However, a detailed examination of amino acid and nucleotide substitutions in sequences of MOMP and *omp1*, respectively, shows no evolutionary pattern with respect to the biological relationships among human strains of *C. trachomatis*.

There is a lack of correlation between disease phenotype and the *omp1* and MOMP phylogenies. Serovars L1, L2, and L3 are responsible for LGV, a disease that is distinct from the urogenital and ocular infections associated with serovars A to K. Yet, there is no pattern in the *omp1* or MOMP phylogenies that groups the LGV serovars distinctly from the urogenital serovars. Likewise, serovars A, B, Ba, and C are most commonly associated with ocular trachoma. While serovars A and C branch within the C-complex clade, they have no specific association with each other. In the nucleotide tree (Fig. 1), they do appear to be distinct from the rest of the C complex, as they are among the first to diverge and have many unique substitutions. However, the substitutions do not appear to be unique to the trachoma biovar since they are not shared by serovars A and C.

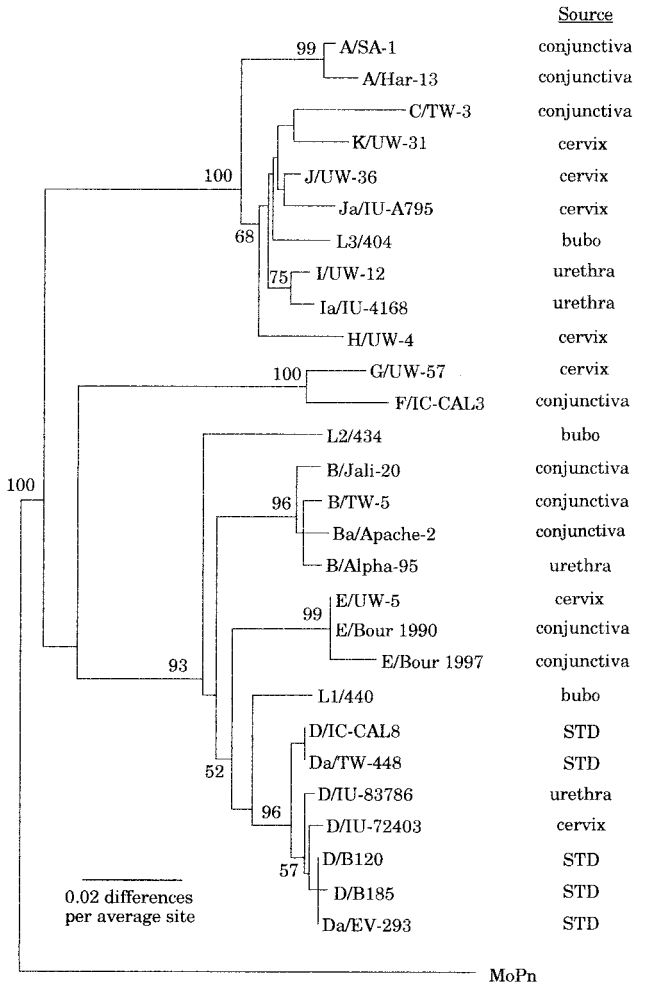

FIG. 2. The distance neighbor-joining tree based on the *C. trachomatis* MOMP amino acid sequence for 29 isolates. The tree was rooted by using the *C. trachomatis* MoPn strain. Branch lengths are proportional to the amount of amino acid sequence divergence between taxa in the tree, as illustrated by the bar. Bootstrap values, based on 1,000 bootstrap replicates, are given at the nodes to which they apply. Bootstrap values below 50 are considered insignificant and are not shown.

If the *omp1* and MOMP phylogenies were reflective of disease, then serovars B and Ba, which are associated with both trachoma and urogenital disease, should be intermediate between the trachoma serovars and the urogenital serovars. However, this is not the case. An examination of ocular and urogenital isolates of serovars Ba and C by Frost et al. (17) showed eight nucleotide and two amino acid differences in conserved segment 1 of MOMP (CS1) between sequences of urogenital Ba isolates (named serovar Bb) and those of ocular Ba isolates. These Bb isolates differed from the prototype D/B120 and H/UW4 in the CS1 by only one base, possibly indicating "urogenital-specific" substitutions in CS1. We have also identified this genotype in the present study, for isolate B/IU-1226. It differs from serovar D and H genotypes by one base in CS1, but it is identical to the serovar Ba and B genotypes throughout the rest of the gene. Therefore, if urogenital-specific substitutions are present, they must be located in CS1 and not in a variable region. A more comprehensive study of serovar B ocular and urogenital genotypes would be necessary in order to determine whether a consistent difference in genotype occurs between urogenital and ocular isolates of serovar B.

There is lack of concordance between host cell niche (tissue tropism) and *omp1* genotype. Serovars A to K (trachoma biovar) infect epithelial cells, while serovars L1 to L3 (LGV biovar) infect epithelial and lymphatic tissue. Therefore, their intracellular niches are different. However, this difference is not reflected in the phylogeny of the *omp1* and MOMP. LGV, trachoma, and urogenital serovars are intermixed in both the *omp1* and MOMP phylogenies. This would suggest that there are no unique amino acid residues in the MOMP that have evolved to recognize and/or bind to tissue-specific receptors. What leads an LGV, a trachoma, or a urogenital strain to infect a certain subset of cells cannot be identified in the MOMP. Batteiger et al. (3) and Allen et al. (1) have identified differences in the *omp2*- and *omp3*-encoded membrane proteins between LGV and trachoma strains. Therefore, these proteins could be responsible for some of the biological differences between urogenital/trachoma and the LGV strains.

Our data reveal a lack of correlation between success (as measured by prevalence in the population) and genotype. Serovars D, E, and F are the most common serovars found in cases of urogenital infection; therefore, they are the most successful. However, if MOMP residues are responsible for success and consequent fecundity, it is not reflected in the *omp1* and MOMP phylogenies. While variants have been described for serovars D and F, few are described for serovar E, and the substitutions are frequently synonymous. Herein, we sequenced 19 isolates of serovar E and found only 3 to be variant, containing one nonsynonymous substitution, and the three variants were linked epidemiologically.

Conversely, Dean and Millman (12) found that 16% of their serovar E population was variant, reflected by 11 variant genotypes out of 67 isolates. However, their E/Bour MOMP sequence (E/Bour 1997) also was different from that described by Peterson et al. (E/Bour 1990) (32) by 10 nucleotides corresponding to four amino acids. It also differed from the sequence obtained from our own stock of E/Bour. The source of the differences may be due to the fact that the E/Bour strain has been propagated in culture for many years in different labs and may have accumulated mutations during this time. However, this explanation is unlikely, because it would require a very high rate of mutation in the *omp1* gene in the absence of immune selection. Hayes and coworkers (22) found some flux in the prevalence of alleles observed in a trachoma-endemic population over time, but the overall distribution remained constant. Fitch et al. (16) plotted evolutionary distance against year of isolation and found there was no measurable rate of evolution for human *C. trachomatis* isolates, and their calculations were biased by including only the variable regions. We have investigated the possibility of cell culture-induced mutation by continuously propagating nine clinical isolates for 20 passages in cell culture and then comparing the MOMP sequence of the original isolate to the one obtained after the 20th passage in culture. We found no differences (38a). Therefore, the variation in E/Bour reported by Dean and Millman (12) is not easily explained.

A successful phenotype, as displayed by serovars D, E, and F, may be one in which a less vigorous immune response is elicited by the infecting serovar. In this way, the chlamydial organism could go undetected, untreated, and further disseminated. Few studies on the severity of disease as it relates to specific serovars, have been done, and the results have been contradictory. Dean et al. (11) looked at serovars F and G in lower and upper genital tract infections. They found that F variants produced a more severe disease phenotype than the

prototype F genotype does. Batteiger et al. (4) examined 224 cases of urethritis and cervicitis and found a weak correlation between inflammation and serovars F and G but only in men. Overall, they found no association between inflammation and infecting serovar. Van de Laar et al. (40) examined 275 cases and found less obvious disease in men infected with serovars F and G but no differences in women. Therefore, there is no evidence to suggest that production of a more mild disease is a strategy employed by the more prevalent serovars.

Variation within the *omp1* is expressed as nucleotide substitutions as well as IN/DEL events. IN/DEL events are consistent with serocomplexes: the B-serocomplex genes are smallest, followed by serovar F and G genes, followed by C-serocomplex genes. Therefore, the variation in the *omp1* gene between more closely related serovars occurs by a base substitution mechanism while more distantly related serovars differ by base substitutions and IN/DEL mutations. However, there is no pattern to the IN/DEL events that corresponds to disease or tissue tropism. Therefore, the presence or absence of specific amino acid residues in the corresponding protein sequence does not appear to be responsible for tissue-specific invasiveness, enhanced transmission, or disease phenotype. In addition, there are no disease-specific or tissue-specific amino acid substitutions that are held in common between related serovars. In other words, there are no synapomorphic (derived and shared) characters that are unique and specific to eye isolates or cervix-urethral isolates or LGV isolates. The only synapomorphic mutations are two LGV-specific substitutions in CS1. However, neither mutation leads to an LGV-specific amino acid substitution at the site. Therefore, it is highly unlikely that substitutions have occurred at these two sites in the LGV strain sequences as a result of selective forces that are specific to the LGV phenotype.

There are eight conserved cysteine residues in the MOMP sequence, all or some of which may be involved in disulfide tertiary or quaternary interactions. However, there were exceptions; most notably, the E/Bour 1997 and E variant isolates from San Francisco (12) had a serine residue at position 207, while all other strains including MoPn, had a cysteine. This finding may be relevant in that all of the E variants described in that study, except the E/Bour 1997 strain, are geographically linked and may be epidemiologically linked as well. The lack of the cysteine residue in these variants may lead to a conformational change in the protein that enhances infectivity or transmission of serovar E in the local (San Francisco) population. However, since detailed clinical or epidemiologic information was not provided for these strains (12), the significance of the substitution is unknown.

In conclusion, while MOMP is highly variable in sequence, suggesting the use of an adaptive mode of evolution to escape immune pressures, one of the most successful MOMP serovars, serovar E, does not appear to be highly variable and thus would seem to contradict the use of this tactic by *C. trachomatis*. In addition, there is a lack of concordance between *omp1* and MOMP phylogenies and disease phenotype, tissue tropism, or epidemiological success. These observations suggest that the role of MOMP in infection may be passive and that it may not be the major ligand responsible for directing infection of various human cell types. This does not mean that MOMP cannot induce protective immunity in humans. However, immune responses against MOMP may confer weak and short-lived immunity, perhaps because macromolecules other than MOMP may be fundamental determinants of cell tropism, disease phenotype, and epidemiological fitness.

ACKNOWLEDGMENTS

We thank Byron Batteiger, Barbara Van Der Pol, and James Williams for superb technical, scientific, and editorial assistance.

This work was supported by grant AI-31494 to R.B.J. from the National Institute for Allergy and Infectious Diseases, National Institutes of Health.

REFERENCES

1. **Allen, J. E., M. C. Cerrone, P. R. Beatty, and R. S. Stevens.** 1990. Cysteine-rich outer membrane proteins of *Chlamydia trachomatis* display compensatory sequence changes between biovariants. Mol. Microbiol. **4:**1543–1550.
2. **Baehr, W., Y.-X. Zhang, T. Joseph, H. Su, F. E. Nano, K. D. E. 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.
3. **Batteiger, B. E., W. J. Newhall, and R. B. Jones.** 1985. Differences in outer membrane proteins of the lymphogranuloma venereum and trachoma biovars of *Chlamydia trachomatis*. Infect. Immun. **50:**488–494.
4. **Batteiger, B. E., W. Lennington, W. J. Newhall, B. P. Katz, H. T. Morrison, and R. B. Jones.** 1989. Correlation of infecting serovar and local inflammation in genital chlamydia infections. J. Infect. Dis. **160:**332–336.
5. **Bavoil, P., A. Ohlin, and J. Schachter.** 1984. Role of disulfide bonding in outer membrane structure and permeability in *Chlamydia trachomatis*. Infect. Immun. **44:**479–485.
6. **Cabot, E. L., and A. T. Beckenbach.** 1989. Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. Comput. Appl. Biosci. **5:**233–234.
7. **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.
8. **Caldwell, H. D., and R. C. Judd.** 1982. Structural analysis of chlamydial major outer membrane proteins. Infect. Immun. **38:**960–968.
9. **Caldwell, H. D., and J. Schachter.** 1982. Antigenic analysis of the major outer membrane protein of *Chlamydia* spp. Infect. Immun. **35:**1024–1031.
10. **Collett, B. A., W. J. Newhall, R. A. Jersild, Jr., and R. B. Jones.** 1989. Detection of surface-exposed epitopes on *Chlamydia trachomatis* by immune microscopy. J. Gen. Microbiol. **135:**85–94.
11. **Dean, D., E. Oudens, 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.
12. **Dean, D., and K. Millman.** 1997. Molecular and mutation trends analyses of *omp1* alleles for serovar E of *Chlamydia trachomatis*. J. Clin. Invest. **99:**475–483.
13. **Farencena, A., M. Comanducci, M. Donati, G. Ratti, and R. Cevenini.** 1997. Characterization of a new isolate of *Chlamydia trachomatis* which lacks the common plasmid and has properties of biovar trachoma. Infect. Immun. **65:**2965–2969.
14. **Felsenstein, J.** 1989. PHYLIP: phylogeny inference package (version 3.5). Cladistics **5:**164–166.
15. **Fielder, T. J., S. Pal, E. M. Peterson, and L. M. de la Maza.** 1991. Sequence of the gene encoding the major outer membrane protein of the mouse pneumonitis biovar of *Chlamydia trachomatis*. Gene **106:**137–138.
16. **Fitch, W. M., E. M. Peterson, and L. M. de la Maza.** 1993. Phylogenetic analysis of the major outer membrane protein genes of *Chlamydiae*, and its implication for vaccine development. Mol. Biol. Evol. **10:**892–913.
17. **Frost, E. H., S. Deslandes, D. Gendron, D. Bourgaux-Ramoisy, and P. Bopurgaux.** 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.
18. **Hamilton, P. T., and D. P. Malinowski.** 1989. Nucleotide sequence of the major outer membrane protein gene from *Chlamydia trachomatis* serovar H. Nucleic Acids Res. **17:**8366.
19. **Hatch, T. P., D. W. Vance, and E. Al-Hossainy.** 1981. Identification of a major envelope protein in *Chlamydia* spp. J. Bacteriol. **146:**426–429.
20. **Hayes, L. J., and I. N. Clark.** 1990. Nucleotide sequence of the major outer membrane protein gene of *Chlamydia trachomatis* strain A/SA1/OT. Nucleic Acids Res. **18:**6136.
21. **Hayes, L. J., M. A. Pickett, J. W. Conlan, S. Ferris, J. S. Everson, M. E. Ward, and I. N. Clark.** 1990. The major outer membrane proteins of *Chlamydia trachomatis* serovars A and B: intra-serovar amino acid changes do not alter specificities of serovar- and C subspecies-reactive antibody-binding domains. J. Gen. Microbiol. **136:**1559–1566.
22. **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.
23. **Jones, R. B., B. P. Katz, B. Van Der Pol, V. A. Caine, B. E. Batteiger, and W. J. Newhall V.** 1986. Effect of blind passage and multiple samplings on recovery of *Chlamydia trachomatis* from urogenital specimens. J. Clin. Microbiol. **24:**1029–1033.

24. **Jones R. B., B. Van Der Pol, and B. P. Katz.** 1989. Effect of differences in specimen processing and passage technique on recovery of *Chlamydia trachomatis*. J. Clin. Microbiol. **27:**894–898.
25. **Kuo, C. C., N. Takahashi, A. F. Swanson, Y. Ozeki, and S. I. Hakomori.** 1996. An N-linked high-mannose type oligosaccharide, expressed at the major outer membrane protein of *Chlamydia trachomatis*, mediates attachment and infectivity of the microorganism to HeLa cells. J. Clin. Invest. **98:**2813–2818.
26. **Lampe, M. F., K. G. Wong, L. M. Kuehl, and W. E. Stamm.** 1997. *Chlamydia trachomatis* major outer membrane protein variants escape neutralization by both monoclonal antibodies and human immune sera. Infect. Immun. **65:** 317–319.
27. **Morrison, R. P., D. D. Manning, and H. D. Caldwell.** 1992. Immunology of *Chlamydia trachomatis* infections: immunoprotective and immunopathogenic responses, p. 57–84. *In* T. C. Quinn (ed.), Sexually transmitted diseases. Raven Press, New York, N.Y.
28. **Newhall, W. J., and R. B. Jones.** 1983. Disulfide-linked oligomers of the major outer membrane protein of *Chlamydiae*. J. Bacteriol. **154:**998–1001.
29. **Newhall, W. J., P. Terho, C. E. Wilde III, B. E. Batteiger, and R. B. Jones.** 1986. Serovar determination of *Chlamydia trachomatis* isolates by using type-specific monoclonal antibodies. J. Clin. Microbiol. **23:**333–338.
30. **Newhall, W. J.** 1987. Biosynthesis and disulfide cross-linking of outer membrane components during the growth cycle of *Chlamydia trachomatis*. Infect. Immun. **55:**162–168.
31. **Peeling, R. I., I. W. Maclean, and R. C. Brunham.** 1984. In vitro neutralization of *Chlamydia trachomatis* with monoclonal antibodies to an epitope on the major outer membrane protein. Infect. Immun. **46:**484–489.
32. **Peterson, E. M., B. A. Markoff and L. M. de la Maza.** 1990. The major outer membrane protein nucleotide sequence of *Chlamydia trachomatis*, serovar E. Nucleic Acids Res. **18:**3414.
33. **Pickett, M. A., M. E. Ward, and I. N. Clarke.** 1987. Complete nucleotide sequence of the major outer membrane protein gene from *Chlamydia trachomatis* serovar L1. FEMS Microbiol. Lett. **42:**185–190.
34. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
35. **Sayada, C., E. Denamur, and J. Elion.** 1992. Complete sequence of the major outer membrane protein-encoding gene of *Chlamydia trachomatis* serovar Da*. Gene **120:**129–130.
36. **Sayada, C., E. Vretou, J. Orfila, J. Elion, and E. Denamur.** 1995. Heterogeneity within the first constant segment of the major outer membrane protein gene in *Chlamydia trachomatis* serovar D/Da distinguishes two lineages. C. R. Acad. Sci. Paris **318:**943–949.
37. **Stephens, R. S., G. T. Mullenbach, R. Sanchez-Pescador, and N. Agabian.** 1986. Sequence analysis of the major outer membrane protein gene from *Chlamydia trachomatis* serovar L2. J. Bacteriol. **168:**1277–1282.
38. **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.
38a. **Stothard, D. R., B. Van Der Pol, J. Williams, N. Cowley, and R. B. Jones.** Unpublished data.
39. **Swanson, A. F., and C.-C. Kuo.** 1991. Evidence that the major outer membrane protein of *Chlamydia trachomatis* is glycosylated. Infect. Immun. **59:** 2120–2125.
40. **Van De Laar, M. J. W., J. Lan, Y. T. H. P. Van Duynhoven, J. S. A. Fennema, J. M. Ossewaarde, A. J. C. Van Den Brule, G. J. J. Van Doornum, R. A. Coutirho, and J. A. R. Van Den Hoek.** 1996. Differences in clinical manifestations of genital chlamydial infections related to serovars. Genitourin. Med. **72:**261–265.
41. **Wang, S. P., C.-C. Kuo, and J. T. Grayston.** 1973. A simplified method for immunological typing of trachoma-inclusion conjunctivitis-lymphogranuloma venereum organisms. Infect. Immun. **8:**356–360.
42. **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.
43. **Wang, S. P., and J. T. Grayston.** 1991. Three new serovars of *Chlamydia trachomatis*: Da, Ia, and L2a. J. Infect. Dis. **163:**403–405.
44. **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.
45. **Zhang, Y.-X., S. Stewart, T. Joseph, H. R. Taylor, and H. D. Caldwell.** 1987. Protective monoclonal antibodies recognize epitopes located on the major outer membrane protein of *Chlamydia trachomatis*. J. Immunol. **138;**575–581.
46. **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.
47. **Zhang, Y.-X., S. G. Morrison, and H. D. Caldwell.** 1990. The nucleotide sequence of the major outer membrane protein gene of *Chlamydia trachomatis* serovar F. Nucleic Acids Res. **18:**1061.

Editor: J. G. Cannon