# GENETICS

# Identification of Serovar-Specific Single-Nucleotide Polymorphisms of *C. Trachomatis omp1* Gene

**V. I. Sharkeev, I. Yu. Torshin, M. M. Shkarupeta, T. A. Akopian, V. N. Lazarev, and V. M. Govorun**

Translated from *Byulleten' Eksperimental'noi Biologii i Meditsiny*, Vol. 140, No. 8, pp. 192-197, August, 2005
Original article submitted December 22, 2004

Complete sequences of *omp1* gene encoding chlamydial main outer membrane protein were analyzed in 76 clinical strains of *C. trachomatis*. Thirty-four serovar-specific single-nucleotide polymorphisms were identified, 20 of them meet two criteria: unique position of the nucleotide and unique nucleotide substitution. Evaluation of serovar-specific single-nucleotide polymorphisms of *omp1* gene can appreciably simplify and accelerate genetic diagnosis of *C. trachomatis* serovars.

**Key Words:** *C. trachomatis omp1 gene; serovars; polymorphisms*

*C. trachomatis* is a gram-negative microorganism, obligate intracellular parasite causing more than 20 nosological entities of human diseases. Classification of *C. trachomatis* species is based on differences in the antigenic epitopes of the main outer membrane protein (MOMP, 40 kDa), a product of *omp1* gene (length of coding sequence 1182-1194 n. p.) [1,6,13]. A total of 15-19 serovars of *C. trachomatis* are distinguished [6,15]. *C. trachomatis* serovars A, B, Ba, C cause mainly trachoma (trachoma biovar) [1,4,9,15]. *C. trachomatis* serovars L1, L2, L2a, L3 cause venereal lymphogranuloma (LGV biovar) [1,4,9]. *C. trachomatis* serovars E, F, D, J, G, Ia/I, K, B/Ba, H, and Da [14] cause urogenital infections (they are listed in the order of their decreasing incidence), though the data on their virulence and clinical manifestations of infections caused by them are contradictory [2,4,7].

Identification of serovar is essential for correct diagnosis of chlamydial infection and epidemiological studies. Serovar-specific antigenic epitopes of *C. trachomatis* MOMP can be identified using monoclonal antibodies, but serotyping with the use of antibodies is not always accurate. For example, serovars L1 and L3 cannot be differentiated from serovars E and G [13]. Amino acid composition of MOMP and *omp1* gene nucleotide sequence contain four so-called variable VS1-VS4 segments divided by constant sequences [5,9]. Nucleotide sequences in *omp1* gene correspond to MOMP amino acid sequences, and hence, genotyping is a method alternative to serotyping.

Detection of serovar-specific polymorphisms, specifically, single nucleotide polymorphisms (SNP), will appreciably simplify and accelerate genetic diagnosis of serovars.

The aim of this study was to detect all serovar-specific polymorphisms of *C. trachomatis omp1* gene. Published (Genbank) and identified in our laboratory *omp1* gene sequences of *C. trachomatis* strains and clinical isolates were used in the study.

## MATERIALS AND METHODS

Clinical strains of *C. trachomatis* were isolated from epithelial scrapings of the mucosa of cervical canal or urethra by inoculation in McCoy cell cultures. *C. trachomatis* DNA was isolated from cell culture by the

Institute of Physicochemical Medicine, Ministry of Health of Russian Federation, Moscow. ***Address for correspondence***: lazar0@newmail.ru. V. N. Lazarev

0007-4888/05/1402-0222 

**TABLE 1.** *C. trachomatis* Strains and Clinical Isolates

| Serovar | Strain/clinical isolate | Length, n. p. | Year of isolation | GenBank No. |
|---|---|---|---|---|
| A | A/Har-13 | 1191 | 1958 | J03813 |
| | A/Sa-1/OT | 1191 | 1957 | M58938 |
| B | B/alpha-95 | 1185 | 1996 | U80075 |
| | B/Jali-20 | 1185 | 1985 | M33636 |
| | B/TW-5 | 1185 | 2000 | AF304856 |
| | B/TW-5/OT | 1185 | 1959 | M17342 |
| | B-099 | 1014 | 2003 | AY464143 |
| | B-3537534 | 865 | 2003 | AY378284 |
| | B/IU-1226 | 942 | 1987 | AF063208 |
| Ba | Ba/Apache-2 | 1157 | 1960 | AF063194 |
| C | C/TW-3 | 1194 | — | AF202455 |
| | C/TW-3/OT | 1194 | — | AF352789 |
| | CH-4312 | 864 | 2003 | AY380109 |
| D | D/IU-71960 | 1154 | 1987 | AF063195 |
| | D/IU-72403 | 1154 | 1987 | AF063196 |
| | D/IU-83786 | 1154 | 1988 | AF063197 |
| | D/B-120 | 1182 | 1983-1991 | X62918 |
| | D/B-185 | 1182 | 1983-1991 | X62919 |
| | D/IC-Cal-8 | 1182 | 1991 | X62920 |
| | D/UW-3/Cx | 1182 | — | NC 000117 |
| | D/LSU-EP212 | 1182 | 2000 | AF279587 |
| | D/LSU-PM12 | 1182 | 2000 | AF279588 |
| | D/Ep6 | 1008 | | X77364 |
| Da | Da/Ev-293 | 1182 | 1983-1991 | X77365 |
| | Da/TW-448 | 1182 | 1991 | X62921 |
| E | E/Bour-1990 | 1182 | 1959 | X52557 |
| | E/Bour-1997 | 1143 | 1959 | U78763 |
| | E/IU-51538 | 1154 | 1986 | AF063198 |
| | E/1 | 1143 | 1989-1992 | U78528 |
| | E/2 | 1143 | 1989-1992 | U78529 |
| | E/3 | 1143 | 1989-1992 | U78530 |
| | E/15 | 1143 | 1989-1992 | U78531 |
| | E/18 | 1143 | 1989-1992 | U78532 |
| | E/55 | 1143 | 1989-1992 | U78537 |
| | E/67 | 1143 | 1989-1992 | U78538 |
| F | F/IC-Cal-3 | 1188 | 1960 | X52080 |
| | F-449 | 1070 | — | AY464145 |
| | F/V-12 | 986 | — | AF265240 |
| | F/IU-1552 | 945 | 1997 | AF063212 |
| G | G/UW-57/Cx | 1160 | 1971 | AF063199 |
| | G-646 | 1103 | — | AY464158 |
| | G-553 | 1073 | 2003 | AY464156 |
| | G/IS-1558 | 1032 | — | AF414957 |
| H | H/IS-1075 | 990 | — | AF414959 |
| | H/UW-4 | 1194 | 1965 | X16007 |
| | H-563 | 954 | — | AY464146 |

*TABLE 1. (sequential)*

| Serovar | Strain/clinical isolate | Length, n. p. | Year of isolation | GenBank No. |
|---|---|---|---|---|
| I | I/UW-12/Ur | 1191 | 1966 | AF063200 |
| | I/IS-655 | 1017 | — | AF414960 |
| Ia | Ia/UW-202 | 1194 | — | AF202456 |
| | Ia/IU-4168 | 1166 | 1987 | AF063201 |
| | Ia/IU-1558 | 951 | 1997 | AF063205 |
| J/Ja | J/UW-36 | 1194 | — | AF202457 |
| | J/UW-36/Cx | 1194 | 1971 | AF063202 |
| | J-149 | 1047 | — | AY464149 |
| | J/IS-326 | 1015 | — | AF414964 |
| | J/IU-117542 | 1163 | 1985 | AF086856 |
| | Ja/IU-37538 | 1194 | | AF063203 |
| K | K/UW-31/Cx | 1194 | 1973 | AF063204 |
| | K-23 | 992 | — | AF265239 |
| L1 | L1/440 | 1182 | 1968 | M36533 |
| — | LGV-98 | 1182 | 1994 | L35605 |
| | LGV-115 | 1182 | 1994 | L35606 |
| L2 | L2/434/Bu | 1185 | 1968 | M14738 |
| L2a | L2a/UW-396 | 1185 | 1985 | AF304858 |
| — | L2b/AMSTLGV | 949 | — | AY586530 |
| L3 | L3/404 | 1194 | 1967 | X55700 |

**Note.** "—": no data.

method of Boom. Primers for amplification of 123-1172 n. p. encoding sequence of *omp1* gene were selected using Vector NTI 9 software (Informax): 1s primer - 5"AGGTTTCGGCGGAGATCC 3'; 1as primer - 5"ACGTGAGCTGCTCTCTCATCG 3' (strain D/UW-3/Cx, Genbank NC 000117). PCR was carried out in 50 μl reaction mixture containing 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 2 mM $MgCl_2$, 100 μM each dNTP (Promega), 0.5 μM each primer, 1 U *Taq* polymerase (Promega), and 50 ng *C. trachomatis* DNA. The protocol of the reaction in a Gene Amp 2400 programmed thermostat (Perkin Elmer) was as follows: 30 sec at 94°C, 30 sec at 60°C, 1 min at 72°C (25 cycles). The reaction products were separated and identified in 1% agarose gel prepared on Tris-acetate buffer (50 mM Tris-HCl (pH 8.0), 50 mM $CH_3COONa$, 5 mM EDTA), stained with ethidium bromide (0.5 μg/ml). Amplification product was purified using Wizard PCR Preps DNA Purification System (Promega) according to manufacturer's instructions. Amplicon nucleotide sequence was determined using Big DyeTM Terminator 3.0 Cycle Sequencing kit (Applied Biosystems in accordance with manufacturer's instruction.

Analysis of nucleotide sequences of encoding region of *omp1* gene and amino acid sequences of MOMP was carried out for 66 *C. trachomatis* strains and clinical isolates from GenBank belonging to 18 serovars (Table 1) and for 10 clinical isolates from our laboratory. Sequences from GenBank belong to strains/isolates collected in 1958-2003. We found that in addition to known *C. trachomatis* serovars [12,15], reports about new serological variants, which cannot be classified using the known schemes, appear from time to time. For example, Ga and Ja variants are described by different authors as "serologic" and as "genetic" [3,6]. Therefore, in our study we used only serotyped serovars.

Nucleotide and amino acid sequences of *omp1* gene were compared using ClustalW software [11]. Since insert/deletion sites are not serovar-specific markers [9], the main task was to identify oligonucleotide polymorphisms. Only polymorphisms detected in all variants of sequences of a serovar were considered as serovar-specific.

Numeration of nucleotide polymorphisms and amino acid variants of MOMP corresponded to encoding region of *omp1* gene of D/UW-3/Cx strain, because complete genome was sequenced for this strain. Evaluation of appurtenance to this or that *C. trachomatis* serovar on the basis of serovar-specific polymorphisms was tested on 21 clinical isolates from GenBank [10] and 10 clinical isolates obtained in our laboratory.

## RESULTS

Variations in *omp1* gene composition are "inserts" and single nucleotide "substitutions" (inserts and substitutions relatively to serovar D *omp1* gene sequence). None of three detected inserts (two-codone 282NNNN NN283 and single-codone 510NNN511 and 996NN N997, numeration for serovar D) was serovar-specific.

All 34 serovar-specific polymorphisms were single-nucleotide (SNP). Of all detected serovar-specific SNP, no more than 20, conforming to two criteria (unique position of nucleotide and unique nucleotide substitution) were sufficient for genotyping (Table 2). For example, of 4 SNP-specific polymorphisms for serovar A, only G298 conformed to these criteria. Hence, of 3 polymorphisms for serovars B/Ba it is sufficient to genotype only G468.

Serovar E, most epidemiologically prevalent, contains seven nucleotide polymorphisms. All of them are located in the third constant sequence. Three of these "silent" polymorphisms are unique (for serovar D MOMP amino acid sequence): C558 (Tyr186), C573 (Phe191), G609 (Glu203); 2 lead to amino acid substitution C612 (Cys204→Ser) and T574(Ala192→Ser).

Serovar-specific polymorphisms in encoding region of *omp1* gene were identified for serovars A, B/Ba, C, D/Da, E, F, G, H, K, and biovar LGV. No serovar-specific polymorphisms for serovars I/Ia and J/Ja were found.

The possibility of biovar-specific polymorphisms was also analyzed. Of three *C. trachomatis* biovars, biovar-specific polymorphisms were detected only for biovar LGV (group of serovars causing venereal lymphogranuloma): G30 (Val10) and A117 (Leu39). No biovar-specific polymorphisms were detected for trachoma biovar (serovars A-C) and urogenital infection biovar (D-K). However, serovars can be identified with high accuracy using the detected polymorphisms (Table 1) and subsequent referring of a serovar to an appropriate biovar.

Most polymorphisms are located in the so-called variable sequences of *omp1* gene, mainly in the first and second sequences. On the other hand, all nucleotide polymorphisms specific for serovar E are located in the third constant segment of *omp1* gene. The notions of "variable" and "constant" segments are inaccurate, because only the use of SNP located in the constant segment helps to discriminate between the sequences of serovar E.

The proposed algorithm of serotyping was used with partial sequences of *omp1* gene from 21 *C. trachomatis* clinical isolates [10]. Each of these sequences was presented by only variable segments 1-4 (Genbank Nos. AF178221-178304). Hence, serovar-specific SNP were detected for only 9 isolates, because SNP for other isolates lay beyond the variable segments. Other 12 isolates presumably belong to serovar E.

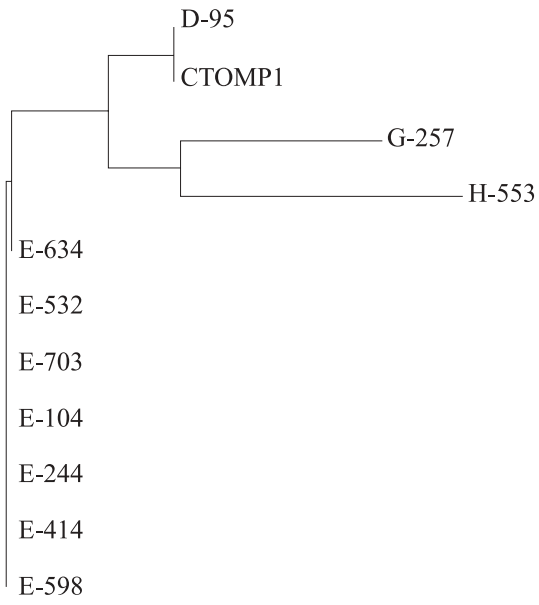


**Fig. 1.** Philogenetic analysis of *omp1* gene sequences and identification of genotype for 10 clinical isolates of *C. trachomatis*. CTOMP1: *omp1* sequence from serovar D complete genome.

The isolates were not serotyped [10]. Their appurtenance to serovars was identified by the authors only by comparison with the reference sequences of different serovars, which the authors do not present in their report. Partial sequences previously determined for one isolate of serovar K (CA13) [10] were very short and did not include serovar-specific polymorphism T734 (Val245). No serovar-specific single-nucleotide polymorphisms were found for serovar Ia, and therefore isolate CA41 was erroneously (by the exclusion method) identified as belonging to serovar E. In general, the proposed SNP correctly identified 19 of 21 isolates. Erroneous identification of these two isolates (CA13 and CA41) was due to incompleteness of the sequences presented previously [10], but not to the deficiency of SNP algorithm proposed in this study.

Sequences of *omp1* gene from 10 clinical isolates of *C. trachomatis*, identified in our laboratory, were analyzed using BLAST software and SNP algorithm developed by the authors. The use of SNP algorithm gives results completely coinciding with the results of identification by BLAST software (Fig. 1). Seven of ten isolates were referred to serovar E and others to serovars D, G, and H (one isolate in each serovar). Hence, analysis of several nucleotides in specific positions distinctly differentiates between sequences belonging to different serovars.

Thus, we propose an algorithm for genotyping *C. trachomatis* intraspecies differences. In comparison with whole gene sequencing, the advantages of SNP

**TABLE 2.** *C. trachomatis omp1* Gene Serovar-Specific Polymorphisms

| Serovar | SNP distribution by gene segments | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1st constant | 1st variable | 2nd constant | 2nd variable | 3rd constant | 3rd variable | 4th constant | 4th variable | 5th constant |
| A | | G288 (K 96)<br>G295 (V 99)<br>T296 (V 99)<br>*G298 (V 99) | | | | | | | |
| B, Ba | | | G468 (V156) | | | G741 (L 247) | | | |
| C | | C303 (N 101) | | *G538 (E 180) | | *A748 (N 250) | | | |
| D, Da | | G284 (S 95) | | A497 (K 166)<br>A498 (K 166) | | | | | |
| E | | | | | *C558 (Y 186)<br>G585 (V 195)<br>*C573 (F 191)<br>*T574 (S 192)<br>A594 (R 198)<br>*G609 (G 203)<br>*C612 (S 204) | | | | |
| F | | *A376 (T 126) | | T 488 (V 163) | *A664 (I 222) | | | *G923 (R 308)<br>G992 (G 331)<br>*T929 (V 310) | |
| G | | | | C499 (Q 167) | | | *A898 (N 300) | | *A1044 (L 348) |
| H | | *A284 (D 95)<br>C/A296(K 99) | | | | | | | |
| K | | | | | *C567(T189) | T734 (V 245) | | | |
| LGV-biovar | *G30 (V 10)<br>*A117 (L 39) | | | | | | | | |
| Total number of SNP | 2 | 9 | 1 | 5 | 9 | 3 | 1 | 3 | 1 |

**Note.** Numeration by D/UW-3/Cx genotype (from complete sequence of *C. trachomatis* genome). *Nucleotide substitutions unique for this serovar in this position.

analysis are obvious, because identification of serovar-specific single-nucleotide polymorphisms markedly simplifies and accelerates genetic diagnosis of serovars. In addition, this approach is indispensable for typing and proving the isolation of a mixture of *C. trachomatis* serovars from patients.

## REFERENCES

1. T. Y. Choi, D. A. Kim, and Y. H. Seo, *J. Korean Med. Sci.*, **16**, No. 1, 15-19 (2001).
2. D. Dean, E. Oudens, G. Bolan, *et al.*, *J. Infect. Dis.*, **172**, No. 4, 1013-1022 (1995).
3. D. Dean, R. J. Suchland, and W. E. Stamm, *Ibid.*, **182**, No. 3, 909-916 (2000).
4. J. I. Ito Jr., J. M. Lyons, and L. P. Airo-Brown, *Infect. Immun.*, **58**, No. 6, 2021-2023 (1990).
5. K. L. Millman, S. Tavare, and D. Dean, *J. Bacteriol.*, **183**, No. 20, 5997-6008 (2001).
6. S. A. Morre, J. M. Ossewaarde, J. Lan, *et al.*, *J. Clin. Microbiol.*, **36**, No. 2, 345-351 (1998).
7. S. A. Morre, L. Rozendaal, I. G. van Valkengoed, *et al.*, *Ibid.*, **38**, No. 6, 2292-2296 (2000).
8. E. J. Postema, L. Remeijer, and W. I. van der Meijden, *Genitourin. Med.*, **72**, 203-205 (1996).
9. D. R. Stothard, G. Boguslawski, and R. B. Jones, *Infect. Immun.*, **66**, No. 8, 3618-3625 (1998).
10. K. Sturm-Ramirez, H. Brumblay, K. Diop, *et al.*, *J. Clin. Microbiol.*, **38**, No. 1, 138-145 (2000).
11. J. D. Thompson, D. G. Higgins, and T. J. Gibson, *Nucleic Acids Res.*, **22**, No. 22, 4673-4680 (1994).
12. S. P. Wang and J. T. Grayston, *J. Clin. Microbiol.*, **29**, No. 7, 1295-1298 (1991).
13. S. P. Wang, C. C. Kuo, R. C. Barnes, *et al.*, *J. Infect. Dis.*, **152**, No. 4, 791-800 (1985).
14. Y. Yuan, Y. X. Zhang, N. G. Watkins, and H. D. Caldwell, *Infect. Immun.*, **57**, No. 4, 1040-1049 (1989).
15. Y. X. Zhang, S. J. Stewart, and H. D. Caldwell, *Ibid.*, No. 2, 636-638.