**Predicting Phenotype and Emerging Strains among Chlamydia trachomatis Infections**

Deborah Dean, William J. Bruno, Raymond Wan, João P. Gomes, Stéphanie Devignot, Tigist Mehari, Henry J.C. de Vries, Servaas A. Morré, Garry Myers, Timothy D. Read, and Brian G. Spratt

*Chlamydia trachomatis* is a global cause of blinding trachoma and sexually transmitted infections (STIs). We used comparative genomics of the family Chlamydiaceae to select conserved housekeeping genes for *C. trachomatis* multilocus sequencing, characterizing 19 reference and 68 clinical isolates from 6 continental/subcontinental regions. There were 44 sequence types (ST). Identical STs for STI isolates were recovered from different regions, whereas STs for trachoma isolates were restricted by continent. Twenty-nine of 52 alleles had nonuniform distributions of frequencies across regions (p<0.001). Phylogenetic analysis showed 3 disease clusters: invasive lymphogranuloma venereum strains, globally prevalent noninvasive STI strains (*ompA* genotypes D/Da, E, and F), and nonprevalent STI strains with a trachoma subcluster. Recombinant strains were observed among STI clusters. Single nucleotide polymorphisms (SNPs) were predictive of disease specificity. Multilocus and SNP typing can now be used to detect diverse and emerging *C. trachomatis* strains for epidemiologic and evolutionary studies of trachoma and STI populations worldwide.

*Chlamydia trachomatis* is a global cause of blinding trachoma and sexually transmitted infections (STIs). Various typing techniques have been developed to better understand the epidemiology and pathogenesis of chlamydial diseases. Early typing schemes used monoclonal and polyclonal antibodies directed against the major outer membrane (MOMP) (1), and differentiated the organism into serovars and seroclasses: B class (comprising serovars B, Ba, D, Da, E, L1, L2a), C class (A, C, H, I, Ia, J, Ja, K, L1, L3), and intermediate class (F, G). Sequencing of *ompA*, which encodes MOMP, has refined typing, detecting numerous trachoma (2,3) and sexually transmitted (4,5) serovar subtypes.

Seroclasses, however, do not correlate with disease phenotypes. For example, A, B, Ba, and C are responsible for trachoma, whereas lymphogranuloma venereum (LGV) strains, L1-3, are associated with invasive diseases, such as suppurative lymphadenitis and hemorrhagic proctitis (6). Other typing techniques, such as restriction fragment length polymorphism (7), random amplification of polymorphic DNA, or pulsed-field gel electrophoresis (PFGE) (8), and amplified fragment length polymorphism (9) also correlate poorly with disease phenotype, and none have been standardized across laboratories.

Multilocus sequence typing (MLST) has been used to characterize strains and lineages of numerous pathogens associated with human diseases that cause serious illness and death, including *Neisseria meningitidis*, *Staphylococcus aureus*, *Vibrio cholerae*, and *Haemophilus influenzae*. MLST uses 500–700 bp sequences of internal regions of 6–8 housekeeping genes, excluding genes suspected to be under immune selection (where there is positive selection for sequence diversity) and ribosomal RNA genes (which are multicycopy and too conserved) (10). Advantages of MLST include its precision, allowing simple interlaboratory comparisons, good discrimination between strains, and buffering against the distorting effect of recombination on genetic relatedness. MLST data are also amenable to various population genetic analyses (11,12). Databases for >30 species are curated at www.mlst.net and pubmlst.org.
parallel with our study, 2 multilocus schemes have recently been developed for *C. trachomatis*. The first violated the above premise by including *ompA*, which is under immune selection (13). The second included only laboratory-adapted and 5 clinical E strains from the Netherlands (14).

In this work, unlike the other *C. trachomatis* MLST schemes, we used complete genomic comparisons of 7 strains from 4 species within the family *Chlamydiaceae* to identify conserved candidate housekeeping genes across the genomes. This approach ensures that the chosen loci are stable over the course of evolution, and allows for future application of a unified MLST scheme for other *Chlamydiaceae* spp. We typed a diverse worldwide collection of reference and clinical isolates from trachoma and STI populations, correlating genetic variation with geography and disease phenotype. We found disease-specific single nucleotide polymorphisms (SNPs) and a diversity of new strains including recombinant strains that occurred for *ompA* relative to housekeeping loci, following up on our recent discovery of this phenomenon at multiple loci in *Chlamydiaceae* genomes (15–17).

**Methods**

**Reference and Clinical Samples**

Nineteen *C. trachomatis* reference strains (A/SA-1, B/TW-5, Ba/Apache-2, C/TW-3, D/UW-3, Da/TW-448, E/Bour, F/IC-Cal3, G/UW-57, H/UW-4, I/UW-12, Ia/IU-4168, J/UW-36, Jc/UW-92, K/UW-31, L1/440, L2/434, L2a/TW-396, and L3/404) and 68 clinical isolates from 6 geographic locations worldwide (obtained from patients with trachoma and STIs including proctitis) were analyzed. Because de-identified clinical data and samples were used, the research was considered institutional review board exempt by Children’s Hospital Oakland Research Institute.

**Selection of Housekeeping Genes**

We genome-sequenced 7 strains from 4 species of the 2 genera of *Chlamydiaceae*: *C. trachomatis* (strains D/UW-3/CX [18] and A/Har-13 [19]), *Chlamydia muridarum* (rodent strain MoPn [20]), *Chlamydonphila pneumoniae* (human strains AR39 [20]; CWL029 [21], and J183 [21]), and *Chlamydonphila caviae* (guinea pig inclusion conjunctivitis strain [22]), the most distantly related species of *Chlamydiaceae*. On the basis of comparative genomics [20] and comparisons generated by CGView [23], we identified an initial candidate pool of 14 housekeeping genes (Figure 1) present in all 7 genomes with an average BLAST score ratio (BSR) (24) $>0.5$ for orthologs queried against *C. caviae* relative to the BLAST score of each sequence against itself. The BSR of $>0.5$ provides a cutoff to select genes that have lower levels of nucleotide sequence divergence in the genome (i.e., putative housekeeping genes). We then selected 7 genes (Figure 1) on the basis of i) diverse chromosomal regions where a single recombinational exchange would be unlikely to co-introduce $>1$ selected gene; ii) regions where several contiguous genes were involved in metabolic or key functions; iii) essential metabolic enzymes (e.g., tRNA synthases); iv) genes without similarity to human genes; and v) no genes under diversifying selection.

**ompA and MLST Analyses**

DNA was extracted from isolates using Roche High Pure Kits (Roche Diagnostics, Pleasanton, CA, USA) and *ompA* genotyped as described (15,16,25). DNASTAR (Madison, WI, USA) was used to design primers to amplify $≈600–700$ bp for each gene (Table 1); BLAST (NCBI; http://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to ensure primer specificity for *C. trachomatis* genes. MLST PCR was carried out in 96-well plates as described (26). Sequenced
Chlamydia trachomatis

Infections

DNA (GenBank accession nos. FJ45414–FJ746022) using ABI3700 instruments were aligned by using MegAlign (DNASTAR). Each unique sequence for a locus was designated as a unique allele using Sequence Output (www.mlst.net). Each allelic profile (made up of the string of integers corresponding to allele numbers at the 7 loci) was assigned as a different strain or clone and given an ST as a clone descriptor. All STs have been deposited in the C. trachomatis site at www.mlst.net.

Allelic profiles and concatenated sequences were used to determine the relatedness of isolates. Average pairwise diversity between isolates was calculated from the 3,714-bp concatenated sequence of the 7 loci for each isolate joined in-frame using MEGA4 (27). Synonymous (dS) and nonsynonymous (dN) substitutions were determined using MEGA4 for each locus. Allele frequencies per locus and geographic region were calculated using SAS software 9.2 (SAS Institute, Inc., Cary, NC, USA) with the PROC FREQ tool supplying the frequency count. We calculated a classification index (11) on the basis of allele and ST frequency between populations of different geographic regions to determine the probability of association of an allele with a particular continental/subcontinent region. Statistical significance was determined by 10,000 resamplings of allele and ST frequency per region.

Strain Clustering and Single Nucleotide Polymorphism Analyses

eBURST (www.mlst.net) was used to identify clusters of related and singleton STs that were not closely related to any other ST (12) and to predict patterns of evolutionary descent. MEGA4 (27) was used to construct a tree from concatenated sequences by using minimum evolution, neighbor joining, or unweighted pair group method with arithmetic mean, with various substitution models including Kimura 2-parameter, Jukes Cantor, and p-distance; 1,000 bootstrap replicates were used to test support for each node in the tree. The short evolutionary distances (<≈0.01) imply that back-substitutions were rare, and as expected, all methods gave similar results (data not shown). SplitsTree (www.splitstree.org) was used for evolutionary tree construction by decomposition analyses using the distance matrix produced from pairwise comparisons of concatenated sequences to determine interconnected networks (28).

A matrix of all SNPs by ST was produced in Excel. SAS was used to identify which SNPs were associated with an ST using PROC FREQ. Statistical significance was determined by using a classification index as above for the probability of association of a SNP with a particular ST. Levene’s test (29) was used to determine whether there was equal variance across the 87 isolates. A p value of <0.05 was considered significant.

Results

Discrimination of C. trachomatis by MLST

Figure 1 shows genomic alignments for C. trachomatis (D/UW-3/CX), C. muridarum, C. pneumoniae (AR39), and C. caviae. The C. trachomatis (A/Har-13 and D/UW-3/CX) and C. pneumoniae (AR39, CWL029, and J138) genome sequences were almost identical within species for gene content and could be represented by D/UW-3/CX and AR39, respectively.

ompA genotypes were compared with STs resolved by MLST. The online Technical Appendix (available from www.cdc.gov/EID/content/15/9/1385-Techapp.pdf) shows ompA genotype (first letter of strain ID), ST, assigned alleles, and clinical characteristics for each isolate. There were 44 STs (0.51 ST/isolate) for the 87 isolates. Thirty STs were represented by a single isolate. In some cases, ST1 isolates from diverse geographic regions shared the same ST. For example, ompA genotype E ST1 isolates (ST39) were found in California, USA; Amsterdam, the Netherlands; Ecuador; Lisbon, Portugal; and Tanzania. Similarly, LGV genotypes (ST1) were identified in San Francisco, California, USA; Seattle, Washington, USA; and Amsterdam; 2 clinical L2b

### Table 1. Primer pairs used for PCR of chlamydiaceae species and strains

| Locus | Region | Primer name | Sequence (5' → 3') | Length of sequence, bp |
|-------|--------|-------------|--------------------|------------------------|
| glyA  | CT432  | FglyA       | GAAGACTGTGGCCGCTGTATTATGG | 522 |
|       |        | RglyA       | CTTCCTGAGGATCCTCTGAC | |
| mdhC  | CT376  | FmdhC       | GGAATCTTTTGCGGCTGTATTGT | 519 |
|       |        | RmdhC       | CGATTACTGCACTACCACGACTCT | |
| pdhA  | CT245  | FpdhA       | CTACAGAAAGGCGCGAGTTT | 549 |
|       |        | RpdhA       | CGTTTGYGTGATGTGCTAGT | |
| yhbG  | CT653  | FyhbG       | TCAAGTCAATGCAGGGAATA | 504 |
|       |        | RyhbG       | GATAGTGGTGACGTCAGCAGTAGAT | |
| pykF  | CT332  | FpykF       | ATCTATCGCGCTCTCGTT | 525 |
|       |        | RpykF       | CAGCAATATTAGGAGATA | |
| lysS  | CT781  | FlsyS       | GAAGGAATCGGAGAAGCGATAAT | 576 |
|       |        | RlysS       | ATACGCCCGATACAGGGAACGAAC | |
| leuS  | CT209  | FlleuS      | TCCCTGGTGTGCAGCTCTCAC | 519 |
|       |        | RleuS       | GGGCATCGCAAAAAACGTAAATAGT | |
genotypes (ST33) were restricted to Amsterdam. In contrast, no trachoma isolates from different continents shared the same ST.

ompA genotypes correlated poorly with relatedness between strains by MLST data (online Technical Appendix). Isolates of the same ST had up to 4 different ompA genotypes. For example, ST19 included ompA genotypes D, H, I, and J. For each ompA genotype, 38%–100% belonged to different STs. Different STs with the same ompA genotype were closely related by MLST (e.g., isolates with C and F ompA genotypes); others were not. Isolates of D, E, and Ja ompA genotypes differed at as many as 5 MLST loci.

Allele Characteristics and Localization by Geography

Allele characteristics are shown in Table 2. The number of alleles at each locus varied from 4 to 11. The average pairwise distance and dS and dN are provided. We determined allele frequencies on the basis of continental/subcontinental regions (Table 3). The majority of alleles were observed multiple times. Seventeen were found only once, and 28 were unique to a specific region (Table 3). The range was from 1 allele at the lysS locus for South America to 9 in North America. The highest frequency of a unique allele was 84.6% (leuS allele 7) for Asia, which also had the highest proportion of unique alleles, 6/17 (35.29%). There was a significant nonuniform distribution of alleles at each locus by classification index.

Phylogenetic Grouping of STs by Disease Phenotypes and Evidence for Recombination

eBURST (11,12) generated 3 clonal complexes (CC) (Figure 2): trachoma strains, A, B, Ba, and C (CC-A); noninvasive STIs with low population prevalence (CC-B); and noninvasive, globally prevalent D/Da, E, and F STIs (CC-C). The online Technical Appendix shows eBURST data, including single, double, and triple locus variants (S/D/TLVs).

Relationships between the isolates was further explored by constructing a minimum-evolution tree using MEGA4 (27). These data showed 3 disease clusters (Figure 3). Cluster I comprised noninvasive STIs (eBURST CC-B) and a trachoma subcluster (eBURST CC-A). Cluster II comprised only invasive LGV strains. Cluster III included noninvasive prevalent D/Da, E and F STIs (eBURST CC-C). E581 strain (ST39; cluster III) was isolated from the conjunctiva of a trachoma patient, most likely representing autoinoculation from the urogenital tract, because all other isolates of this ST were from STIs.

Nine isolates did not localize on the MLST tree with strains of the same ompA genotype (Figure 3). Ja41nl and Ja47nl, which were expected to cluster with other J and Ja isolates in cluster I if the genome sequences were similar, were identical by MLST to reference strain F and clinical isolates F8p, F9p, E19e, and E5s in cluster III. Similarly, D83s, which were expected to cluster with other D and Da isolates in cluster III, had the same ST as H40nl, H18s, I22p, and J44n in cluster I; D2s were identical to Ia and Ia57e in cluster I. Additionally, G16p did not cluster with the other G isolates in cluster I. In analyzing locations of incongruence between clinical D and E isolates in cluster I, compared with those in cluster III, the loci that differed were glyA, yhbG, and pykF in which allele assignments were identical, in general, to G, H, I, Ia, J, Ja, and K strains (online Technical Appendix) in cluster I. These were the exact same loci that differed for Ja41nl and Ja47nl in cluster III, compared with other J/Ja isolates in cluster I. Ja26s differed at glyA, mdhC, and yhbG, whereas G16p differed at yhbG, lysS, and leuS. Furthermore, the ompA tree (Figure 4) was incongruent with the MLST tree. We interpret all 9 isolates to be recombinants.

SplitsTree decomposition evaluated alternative evolutionary pathways that might indicate recombination between MLST loci (Figure 5). There was considerable network structure, providing evidence of alternative pathways between strains, which may indicate that recombination has influenced the evolution of housekeeping genes for the C. trachomatis strains.

SNPs Associated with Disease Phenotypes

We identified 61 polymorphic sites among the 7 loci. Multiple SNPs were significantly associated with each of the 3 clusters and disease groups (Table 4). For example, 15 SNPs in yhbG and leuS were 100% specific for all LGV strains in cluster III. Any 1 of these SNPs could be used to identify these strains. SNPs 4, 29, 31, 33, and 34 (together or any 1 alone) were specific for the cluster II STIs. For

| Gene locus | No. alleles | Length, bp | No. polymorphic sites | Average pairwise distance | Average dS | Average dN |
|------------|-------------|------------|-----------------------|--------------------------|------------|------------|
| glyA       | 7           | 522        | 5                     | 0.003                    | 0.0101     | 0.0034     |
| mdhC       | 4           | 519        | 3                     | 0.001                    | 0.0112     | 0.0025     |
| pdhA       | 7           | 549        | 6                     | 0.0003                   | 0.0007     | 0.0030     |
| yhbG       | 8           | 504        | 21                    | 0.01                     | 0.0670     | 0.0026     |
| pykF       | 7           | 525        | 7                     | 0.003                    | 0.0105     | 0.0034     |
| lysS       | 8           | 576        | 9                     | 0.002                    | 0.0093     | 0.0044     |
| leuS       | 11          | 519        | 10                    | 0.003                    | 0.0104     | 0.0055     |
| Overall    | 52          | 3,714      | 61                    | 0.003                    |             |            |
Chlamydia trachomatis Infections

the trachoma subcluster I, unlike for other clusters, only SNP 38 was associated with reference strains B and C and all clinical trachoma strains; SNPs 54, 55 and 57 together (but not any alone) represented all trachoma strains except reference strain A. Based on classification indices and Levene's test, the null hypothesis of a uniform distribution of SNPs was rejected at the respective locus.

Discussion

Accumulating evidence for recombination among Chlamydiaceae in general, and C. trachomatis in particular, has motivated a typing system that provides buffering from the distorting effects of genetic reshuffling that plague systems based on a single locus. We therefore developed an MLST scheme derived from comparative genomics of species within the family Chlamydiaceae to select conserved chromosomally dispersed housekeeping genes. Our scheme showed considerable variability in allelic profiles associated with geographic regions, as well as diverse and recombinant strains. We also identified SNPs that correlated with the 3 C. trachomatis disease groups: invasive LGV diseases, noninvasive urogenital diseases, and trachoma.

Comparative genomics of Chlamydia and Chlamydophila spp. identified 14 conserved housekeeping genes that could be used to extend MLST schemes for these and potentially other Chlamydiaceae spp.. Surprisingly, each gene was located in a different position within the respective genome, indicating a lack of synteny among chromosomes (20) (Figure 1), except for the 2 C. trachomatis and 3 C. pneumoniae strains, which share within species >99% nucleotide sequence identity. This finding suggests that future schemes should select loci to ensure reasonable coverage of the chromosome.

Although there was relatively little sequence diversity in the housekeeping genes, the number of STs (0.51 ST/isolate) was similar to that of other bacterial pathogens. The
None of the loci were identical to ours. In a recent study of the bacterium *Burkholderia pseudomallei* in Australia, there were 0.65 ST/isolate (11,12) with relatively little diversity and few alleles per locus. However, high levels of recombination are believed to shuffle alleles to generate different large numbers of allelic profiles (STs). The extent to which recombination among alleles generates novel STs in *C. trachomatis* is unclear. Although the number of STs per isolate varies, the majority of MLST schemes have been successful for strain discrimination, epidemiologic studies, and evaluation of organism evolution (10). MLST, however, may not be sufficiently discriminatory for some epidemiologic investigations, even with increased loci numbers. This may be the case for LGV strains, although our scheme resolved 2 L2b strains from all other LGV strains.

We found that a number of STs for STI isolates were shared across continents. This finding was particularly evident for those from Amsterdam, Ecuador, Lisbon, and San Francisco, which would be expected given increasing opportunities for global travel and international sexual encounters. Notably, L2b isolates (ST33) from proctitis cases differed at 2 loci from other LGV isolates (ST1) and were restricted to Amsterdam. Although some L2b strains from Amsterdam and San Francisco have historically been similar (30), ST differentiation most likely reflects the emergence of these strains among men who have sex with men.

Not surprisingly, STs for trachoma isolates were restricted to the geographic region of origin where populations travel only locally.

Allele frequencies were assigned on the basis of continental/subcontinental regions (Table 3). Most alleles were observed multiple times, and more than half were region specific. Despite the opportunity for worldwide spread, some strains may be stable within the respective geographic populations. This stability was particularly evident in Africa and Asia, where the frequency of unique alleles was the highest, although this finding also reflects the fact that most isolates were from trachoma populations. As expected, we found, in general, a statistically significant nonuniform distribution of alleles.

Analyses using eBURST and trees constructed in MEGA4 resolved isolates into clonal complexes or clusters.
Both methods identified distinctive groupings of strains by disease phenotypes. STIs caused by less common strains formed an eBURST group (CC-B) but were within Cluster I on the tree together with the trachoma Subcluster I, which was a separate eBURST group (CC-A). A similar clustering pattern to our tree was found by Pannyhoek et al. (14) by using 16 reference and 5 clinical E strains, but they did not distinguish trachoma reference strain B/TW-5 from the LGV group. Our Cluster II included only LGV strains. Cluster III contained the noninvasive globally prevalent D/Da, E, and F strains (eBURST CC-C). This cluster represents efficiently transmitted strains with adaptive fitness in the genital tract.

A number of isolates representing different ompA genotypes shared the same ST, whereas many isolates of the same ompA genotype had different STs (online Technical Appendix). Furthermore, 9 isolates were found outside the expected cluster, suggesting that recombinational replacement at the ompA locus occurs relatively frequently. Accumulating evidence supports frequent recombination among Chlamydiaceae. Initial evidence came from observations of recombination within ompA (4,31) followed by phylogenetic analyses (32), and bioinformatic and statistical analyses for multiple species of the family Chlamydiaceae and C. trachomatis strains (15). Recently, we showed intergenic recombination involving ompA and pmpC, pmpE-I, and frequent recombination throughout the genome with significant hotspots for recombination for recent clinical isolates (16,17). Pannekoek et al. noted incongruence between ompA and fumC sequences (14). Most recombination in our study involved yhbg, glyA, and pykF (online Appendix Table) with incongruence, compared with ompA. Based on C. trachomatis recombinants that have been created in vitro, the estimated size of transferred DNA ranged from 123 kb to 790 kb (33). Although additional recombination may exist in regions that were not sequenced, any gene in our study could be involved in lateral gene exchange with a range of 1,191 bp for a single gene (e.g., ompA), 27 kb (yhbg to ompA) to at least 248 kb (glyA to yhbg), which is consistent with DeMars and Weinfurter (33) and our previous findings (16,17).

Analysis of the 61 SNPs among the 7 loci showed a statistically significant association of specific polymorphisms with each disease cluster (Table 4). A total of 15 SNPs singly or together identified the LGV cluster. Similarly, 5 SNPs identified the prevalent Cluster II D/Da, E, and F strains. Three clinical D and E strains did not contain these SNPs and each appeared to be a recombinant with other STI strains. Only 1 SNP (in pykF) identified all trachoma strains in Subcluster I. Reference trachoma strains A and Ba did not contain this SNP, suggesting that they may not represent circulating strains among present-day populations.

Other studies have associated SNPs or indels in pmp and porB genes with specific disease causing C. trachomatis clades (16,34,35). However, SNPs were not individually analyzed for specific disease associations and the target genes encode surface exposed proteins likely to be under selection for epitope variation to avoid immune system surveillance. A frame-shift mutation in 1 of the tryptophan synthase genes, trpA, was associated with trachoma strains when compared with all others, although some B and C strains lack the entire gene (35). Large deletions in the cytotoxin loci have also been identified that differentiate the 3 disease groups, yet strain B is missing these loci (36). The latter study relied on reference strains, which may limit the use of these deletions for identifying disease-specific groups because clinical isolates may vary in deletion size or location. Additionally, tryptophan synthase genes and cytotoxin loci are located within the 50-kb plasticity zone.
RESEARCH

of the chromosome, a region known for genetic reshuffling (20). The current study differs from those previously mentioned in that it used housekeeping genes that are not under immune selection or in the plasticity zone. Therefore, the SNPs we identified are probably neutral and can be used as reliable markers for disease association. Furthermore, SNPs were based on reference and clinical isolates of multiples of the same strains from 6 geographic regions, representing a broad diversity of this species.

Given the high rates of infection among STI (37) and trachoma populations (38,39), the ability to distinguish LGV and noninvasive urogenital and trachoma strains, including mixed infections, would aid epidemiologists, clinicians, and public healthcare workers worldwide in determining appropriate therapeutic or intervention strategies (40). Our multilocus and SNP typing can now be used to standardize the way an organism is typed; isolates from diverse geographic regions worldwide can be identified and compared; and diverse and emerging C. trachomatis strains can be detected for epidemiologic and evolutionary studies among trachoma and STI populations worldwide.

Acknowledgments

We thank Hugh Taylor for providing trachoma samples from Tanzania.

This work was supported by NIH grants R01 AI059647, R01 AI039499 and R01 EY/AI012219 (D.D.), a European Union grant EU-FP6-LSHG-CT-2007-037637 (S.A.M.), and a Wellcome Trust Grant 030662 (B.G.S.).

Dr Dean is director of Children’s Global Health Initiative and a faculty member at the University of California at San Francisco and the University of California at Berkeley Joint Graduate Group in Bioengineering. She is also a senior scientist in the Center for Immunobiology and Vaccine Development at Children’s Hospital Oakland Research Institute. Her research interests focus on chlamydial pathogenesis and comparative genomics and molecular epidemiology of chlamydial ocular and sexually transmitted diseases.

Table 4. SNPs and combinations of SNPs that correlate 100% with designated cluster and disease phenotype group

| Gene locus | SNP no. | Cluster III | Cluster II | Subcluster † |
|------------|---------|-------------|------------|--------------|
| glyA       | 1–5     |             | 4‡         |              |
| mdhC       | 6–8     |             |            |              |
| pykF       | 36–42   |             | 38¶        |              |
| lysS       | 43–51   |              |            |              |
| leuS       | 52–61   | 52§         | 54         | 55           |
|            |         | 56          | 55         |              |
|            |         | 61          | 57         |              |
| Total no. SNPs | 61 | 15 | 5 | 4 |

*SNP, single nucleotide polymorphism. Each SNP is 100% specific for designated cluster and disease group (e.g., SNP 15 identifies cluster III but all 15 SNPs are specific to cluster III). Cluster III comprises all clinical and reference lymphogranuloma venereum (LGV) strains; cluster II comprises all reference and clinical D, Da, E, F, and recombinant clinical strains except recombinant clinical D2s, D43nl, E87e, and reference Ja strains; subcluster I comprises all reference and clinical trachoma A, B, Ba and C strains except reference strains A and Ba. Disease phenotype groups: invasive LGV strains (cluster III); trachoma A, B, Ba and C strains (subcluster I); and noninvasive globally prevalent STI D/Da, E, and F strains (cluster II). Clinical Ja strains likely acquired a Ja ompA gene by recombination.

†SNPs 54, 55, and 57 together (not independently) are specific for all trachoma strains except reference A strain.

‡p<0.01 for classification index.

§p<0.001 for classification index.

¶p = 0.008 for classification index.
References

1. Wang SP, Grayston JT. Three new serovars of Chlamydia trachomatis: Da, Ia, and L2a. J Infect Dis. 1991;163:403–5.
2. Dean D, Schachter J, Dawson CR, Stephens RS. 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. 1992;166:383–92.
3. Hayes LJ, Bailey RL, Mabey DC, Clarke IN, Pickett MA, Watt PJ, et al. 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. 1992;166:1173–7.
4. Brunham R, Yang C, Maclean I, Kimani J, Maitha G, Plummer F. Chlamydia trachomatis from individuals in a sexually transmitted diseases core group exhibit frequent sequence variation in the major outer membrane protein (omp1) gene. J Clin Invest. 1994;94:458–63. DOI: 10.1172/JCI117347
5. Dean D, Oudens E, Bolan G, Padian N, Schachter J. Major outer membrane protein variants of Chlamydia trachomatis are associated with severe upper genital tract infections and histopathology in San Francisco. J Infect Dis. 1995;172:1013–22.
6. Hamill M, Benn P, Carder C, Copas A, Ward H, Ison C, et al. The clinical manifestations of anorectal infection with lymphogranuloma venereum (LGV) versus non-LGV strains of Chlamydia trachomatis: a case-control study in homosexual men. Int J STD AIDS. 2007;18:472–5. DOI: 10.1258/095646207781147319
7. Frost EH, Deslandes S, Veilleux S, Bourgau-Ramoisy D. Typing Chlamydia trachomatis by detection of restriction fragment length polymorphism in the gene encoding the major outer membrane protein. J Infect Dis. 1991;163:1103–7.
8. Rodriguez P, Allardet-Servent A, de Barbeyrac B, Ramuz M, Bebear JP. Genetic variability among Chlamydia trachomatis clinical strains analyzed by pulsed-field gel electrophoresis. J Clin Microbiol. 1994;32:2921–8.
9. Morré SA, Ossewaarde JM, Savelkoul PH, Stoof J, Meijer CJ, van den Brule AJ. Analysis of genetic heterogeneity in Chlamydia trachomatis clinical isolates of serovars D, E, and F by amplified fragment length polymorphism. J Clin Microbiol. 2000;38:3463–6.
10. Maiden MC. Multilocus sequence typing of bacteria. Annu Rev Microbiol. 2000;60:561–88. DOI: 10.1146/annurev.micro.54.030804.121325
11. Olsen KA, Wilson DJ, Kriz P, McVean G, Maiden MC. The influence of mutation, recombination, population history, and selection on patterns of genetic diversity in Neisseria meningitidis. Mol Biol Evol. 2005;22:562–9. DOI: 10.1093/molbev/msi041
12. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol. 2004;186:1518–30. DOI: 10.1128/JB.186.5.1518-1530.2004
13. Klint M, Fuxelius HH, Goldkuhl RR, Skarin H, Rutemark C, Andersson SG, et al. High-resolution genotyping of Chlamydia trachomatis strains by multilocus sequence analysis. J Clin Microbiol. 2007;45:1410–4. DOI: 10.1128/JCM.02301-06
14. Pannekoek Y, Morelli G, Kusecek B, Morre SA, Ossewaarde JM, Langerak AA, et al. Multi locus sequence typing of Chlamydiae: clonal groupings within the obligate intracellular bacteria Chlamydia trachomatis. BMC Microbiol. 2008;8:42. DOI: 10.1186/1471-2180-8-42
15. Millman KL, Tavare S, Dean D. Recombination in the ompA gene but not the ompB gene of Chlamydia contributes to serovar-specific differences in tissue tropism, immune surveillance, and persistence of the organism. J Bacteriol. 2001;183:5997–6008. DOI: 10.1128/JB.183.20.5997-6008.2001
16. Gomes JP, Nunes A, Bruno WJ, Borrego MJ, Florindo C, Dean D. Polymorphisms in the nine polymorphic membrane proteins of Chlamydia trachomatis across all serovars: evidence for serovar Da recombination and correlation with tissue tropism. J Bacteriol. 2006;188:275–86. DOI: 10.1128/JB.188.1.275-286.2006
17. Gomes JP, Bruno WJ, Nunes A, Santos N, Florindo C, Borrego MJ, et al. Evolution of Chlamydia trachomatis diversity occurs by widespread interstrain recombination involving hotspots. Genome Res. 2007;17:50–60. DOI: 10.1101/gr.567470
18. Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, Aravind L, et al. Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis. Science. 1998;282:754–9. DOI: 10.1126/science.282.5389.754
19. Carlson JH, Porcella SF, McClarty G, Caldwell HD. Comparative genomic analysis of Chlamydia trachomatis oculotropic and genitalotropic strains. Infect Immun. 2005;73:6407–18. DOI: 10.1128/IAI.73.10.6407-6418.2005
20. Read TD, Brunham RC, Shen C, Gill SR, Heidelberg JF, White O, et al. Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39. Nucleic Acids Res. 2000;28:1397–406. DOI: 10.1093/nar/28.6.1397
21. Shriai M, Hirakawa H, Kimoto M, Tabuchi M, Kishi F, Ouchi K, et al. Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA. Nucleic Acids Res. 2000;28:2311–4. DOI: 10.1093/nar/28.12.2311
22. Read TD, Myers GS, Brunham RC, Nelson WC, Paulsen IT, Heidelberg J, et al. Genome sequence of Chlamydia pneumoniae (Chlamydia psittaci GPIC): examining the role of niche-specific genes in the evolution of the Chlamydiae. Nucleic Acids Res. 2003;31:2134–47. DOI: 10.1093/nar/gkg321
23. Grant JR, Stothard P. The CGView Server: a comparative genomics tool for circular genomes. Nucleic Acids Res. 2008;36:W181–4. DOI: 10.1093/nar/gkn179
24. Rasko DA, Myers GS, Ravel J. Visualization of comparative genomic analyses by BLAST score ratio. BMC Bioinformatics. 2005;6:2. DOI: 10.1186/1471-2105-6-2
25. Millman K, Black CM, Johnson RE, Stamm WE, Jones RB, Hook EW, et al. Population-based genetic and evolutionary analysis of Chlamydia trachomatis urogenital strain variation in the United States. J Bacteriol. 2004;186:2457–65. DOI: 10.1128/JB.186.8.2457-2465.2004
26. Meats E, Feil EJ, Stringer S, Cody AJ, Goldstein R, Kroll JS, et al. Characterization of encapsulated and noncapsulated Haemophilus influenzae and determination of phylogenetic relationships by multilocus sequence typing. J Clin Microbiol. 2003;41:1623–36. DOI: 10.1128/JCM.41.4.1623-1636.2003
27. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol. 2007;24:1596–9. DOI: 10.1093/molbev/msm092
28. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006;23:254–67. DOI: 10.1093/molbev/msj030
29. Levene H. Robust tests for the equality of variances. In: Olkin I, editor. Contributions to probability and statistics: essays in honor of Harold Hotelling. Stanford (CA): Stanford University Press; 1960. p. 278–92.
30. Spaargaren J, Schachter J, Moncada J, de Vries HJ, Fennema HS, Pena AS, et al. Slow epidemic of lymphogranuloma venereum L2b strain. Emerg Infect Dis. 2005;11:1787–8.
31. Hayes LJ, Yearsley P, Treharne JD, Ballard RA, Fehler GH, Ward ME. Evidence for naturally occurring recombination in the gene encoding the major outer membrane protein of lymphogranuloma venereum isolates of Chlamydia trachomatis. Infect Immun. 1994;62:5659–63.
32. Fitch WM, Peterson EM, de la Maza LM. Phylogenetic analysis of the outer-membrane-protein genes of chlamydiae, and its implication for vaccine development. Mol Biol Evol. 1993;10:892–913.

33. DeMars R, Weinfurter J. Interstrain gene transfer in Chlamydia trachomatis in vitro: mechanism and significance. J Bacteriol. 2008;190:1605–14. DOI: 10.1128/JB.01592-07

34. Brunelle BW, Sensabaugh GF. The ompA gene in Chlamydia trachomatis differs in phylogeny and rate of evolution from other regions of the genome. Infect Immun. 2006;74:578–85. DOI: 10.1128/IAI.74.1.578-585.2006

35. Caldwell HD, Wood H, Crane D, Bailey R, Jones RB, Mabey D, et al. Polymorphisms in Chlamydia trachomatis tryptophan synthase genes differentiate between genital and ocular isolates. J Clin Invest. 2003;111:1757–69.

36. Carlson JH, Hughes S, Hogan D, Cieplak G, Sturdevant DE, McClarty G, et al. Polymorphisms in the Chlamydia trachomatis cytotoxin locus associated with ocular and genital isolates. Infect Immun. 2004;72:7063–72. DOI: 10.1128/IAI.72.12.7063-7072.2004

37. Fine D, Dicker L, Mosure D, Berman S. Increasing chlamydia positivity in women screened in family planning clinics: do we know why? Sex Transm Dis. 2008;35:47–52. DOI: 10.1097/OLQ.0b013e31813e0c26

38. Dean D, Kandel RP, Adhikari HK, Hessel T. Multiple Chlamydiaceae species in trachoma: implications for disease pathogenesis and control. PLoS Med. 2008;5:e14. DOI: 10.1371/journal.pmed.0050014

39. Somboonna N, Mead S, Liu J, Dean D. Discovering and differentiating new and emerging clonal populations of Chlamydia trachomatis with a novel shotgun cell culture harvest assay. Emerg Infect Dis. 2008;14:445–53. DOI: 10.3201/eid1403.071071

40. McLean CA, Stoner BP, Workowski KA. Treatment of lymphogranuloma venereum. Clin Infect Dis. 2007;44(Suppl 3):S147–52. DOI: 10.1086/511427

Address for correspondence: Deborah Dean, Director, Children’s Global Health Initiative Children’s Hospital Oakland Research Institute, 5700 Martin Luther King Jr Way, Oakland, CA 94609, USA; email: ddean@chori.org
### Predicting Phenotype and Emerging Strains among *Chlamydia trachomatis* Infections

#### Technical Appendix

**Technical Appendix Table.** Sequence types, allelic profiles, and clinical characteristics of reference and clinical isolates*

| Strain ID† | ST | glyA | mdhC | pdhA | yhbG | pykF | lysS | leuS | Region of isolation | Diagnosis                  |
|------------|----|------|------|------|------|------|------|------|---------------------|---------------------------|
| L1/440     | 1  | 01   | 01   | 03   | 08   | 01   | 04   | 11   | California          | LGV                       |
| L2/434     | 1  | "    | "    | "    | "    | "    | "    | "    | California          | LGV                       |
| L2a/UW-396 | 1  | "    | "    | "    | "    | "    | "    | "    | San Francisco       | Proctitis                 |
| L2b/86nl   | 1  | "    | "    | "    | "    | "    | "    | "    | Seattle             | LGV                       |
| L3/404     | 1  | "    | "    | "    | "    | "    | "    | "    | Amsterdam           | Proctitis                 |
| D/84s      | 2  | 02   | 03   | 03   | 06   | 05   | 04   | 03   | California          | LGV                       |
| H/UW-4     | 3  | 03   | 01   | 03   | 06   | 06   | 04   | 03   | San Francisco       | Cervicitis                |
| A/51t      | 4  | 03   | 03   | 01   | 06   | 03   | 07   | 09   | Tanzania            | Trachoma                  |
| H/46nl     | 5  | 03   | 03   | 02   | 06   | 06   | 08   | 03   | Amsterdam           | Cervicitis and vaginal discharge |
| B/TW-5     | 6  | 03   | 03   | 03   | 04   | 03   | 05   | 10   | Taiwan              | Conjunctivitis            |
| I/UW-2     | 7  | 03   | 03   | 03   | 06   | 01   | 04   | 03   | Washington          | Urethritis                |
| K/UW-31    | 8  | 03   | 03   | 03   | 06   | 02   | 04   | 03   | Washington          | Cervicitis                |
| J/UW36     | 9  | 03   | 03   | 03   | 06   | 02   | 08   | 03   | Washington          | Cervicitis                |
| Ja/UW-92   | 9  | "    | "    | "    | "    | "    | "    | "    | Washington          | Cervicitis                |
| B/53t      | 10 | 03   | 03   | 03   | 06   | 03   | 04   | 09   | Tanzania            | Trachoma                  |
| C/TW-3     | 11 | 03   | 03   | 03   | 06   | 03   | 05   | 07   | Taiwan              | Conjunctivitis            |
| A/48t      | 12 | 03   | 03   | 03   | 06   | 03   | 05   | 09   | Tanzania            | Trachoma                  |
| A/59t      | 12 | "    | "    | "    | "    | "    | "    | "    | Tanzania            | Trachoma                  |
| B/50t      | 12 | "    | "    | "    | "    | "    | "    | "    | Tanzania            | Trachoma                  |
| B/60t      | 12 | "    | "    | "    | "    | "    | "    | "    | Tanzania            | Trachoma                  |
| B/61t      | 12 | "    | "    | "    | "    | "    | "    | "    | Tanzania            | Trachoma                  |
| B/62t      | 12 | "    | "    | "    | "    | "    | "    | "    | Tanzania            | Trachoma                  |
| Ba/52t     | 12 | "    | "    | "    | "    | "    | "    | "    | Tanzania            | Trachoma                  |
| C/32n      | 13 | 03   | 03   | 03   | 06   | 03   | 06   | 07   | Nepal               | Trachoma, TS              |
| C/33n      | 13 | "    | "    | "    | "    | "    | "    | "    | Nepal               | Trachoma, TS              |
| G/16s      | 14 | 03   | 03   | 03   | 06   | 04   | 04   | 08   | San Francisco       | Cervicitis                |
| K/42nl     | 15 | 03   | 03   | 03   | 06   | 06   | 01   | 06   | Amsterdam           | Cervicitis and vaginal discharge |
| J/44nl     | 19 | "    | "    | "    | "    | "    | "    | "    | Amsterdam           | Cervicitis and vaginal discharge |
| J/27s      | 16 | 03   | 03   | 03   | 06   | 06   | 01   | 08   | San Francisco       | Cervicitis/urethritis     |
| E/87e      | 17 | 03   | 03   | 03   | 06   | 06   | 02   | 03   | Ecuador             | Cervicitis                |
| Ba/Apache-2| 18 | 03   | 03   | 03   | 06   | 03   | 03   | 09   | Arizona             | Cervicitis                |
| D/83s      | 19 | 03   | 03   | 03   | 06   | 06   | 04   | 03   | San Francisco       | Cervicitis                |
| H/40nl     | 19 | "    | "    | "    | "    | "    | "    | "    | Amsterdam           | Cervicitis                |
| H/18s      | 19 | "    | "    | "    | "    | "    | "    | "    | San Francisco       | Cervicitis/urethritis     |
| I/22p      | 19 | "    | "    | "    | "    | "    | "    | "    | Lisbon              | Cervicitis/urethritis     |
| J/44nl     | 19 | "    | "    | "    | "    | "    | "    | "    | Amsterdam           | Cervicitis                |
| D/43nl     | 20 | 03   | 03   | 03   | 06   | 06   | 04   | 06   | Amsterdam           | Cervicitis and vaginal discharge |
| G/13s      | 21 | 03   | 03   | 03   | 06   | 06   | 04   | 08   | San Francisco       | Proctitis                 |
| G/14s      | 21 | "    | "    | "    | "    | "    | "    | "    | San Francisco       | Proctitis                 |
| A/SA-1     | 22 | 03   | 03   | 03   | 06   | 06   | 05   | 02   | Saudi Arabia        | Conjunctivitis            |
| la/UW-202  | 23 | 03   | 03   | 03   | 06   | 06   | 08   | 03   | Washington          | Cervicitis                |
| D/2s       | 23 | "    | "    | "    | "    | "    | "    | "    | San Francisco       | Cervicitis/urethritis     |
| la/57e     | 23 | "    | "    | "    | "    | "    | "    | "    | Ecuador             | Cervicitis                |
| la/24s     | 24 | 03   | 03   | 03   | 06   | 06   | 08   | 08   | San Francisco       | Cervicitis/urethritis     |
| la/25s     | 24 | "    | "    | "    | "    | "    | "    | "    | San Francisco       | Cervicitis/urethritis     |
| D/B120     | 25 | 03   | 03   | 03   | 06   | 07   | 04   | 01   | Washington          | Cervicitis                |
| G/16p      | 26 | 03   | 03   | 03   | 07   | 06   | 08   | 05   | Lisbon              | Cervicitis/urethritis     |
| G/17p      | 27 | 03   | 03   | 04   | 06   | 06   | 04   | 03   | Lisbon              | Cervicitis/urethritis     |
| Strain ID† | ST | glyA | mdhC | pdhA | yhbG | pykF | lysS | leuS | Region of isolation     | Diagnosis                  |
|------------|----|------|------|------|------|------|------|------|--------------------------|----------------------------|
| Ia/23p     | 28 | 03   | 03   | 04   | 04   | 04   | 04   | 03   | Lisbon                   | Cervicitis/urethritis      |
| H/21p      | 29 | 03   | 03   | 04   | 06   | 06   | 08   | 03   | Lisbon                   | Cervicitis                 |
| G/UW-57    | 30 | 03   | 03   | 05   | 06   | 06   | 04   | 03   | Washington               | Cervicitis                 |
| H/20p      | 31 | 03   | 03   | 07   | 06   | 06   | 08   | 03   | Lisbon                   | Cervicitis/urethritis      |
| F/38nl     | 32 | 04   | 04   | 03   | 02   | 07   | 04   | 03   | Amsterdam                | Cervicitis and vaginal discharge |
| L2b/48nl   | 33 | 05   | 02   | 03   | 08   | 01   | 04   | 11   | Amsterdam                | Proctitis                  |
| L7/b55nl   | 33 | "    | "    | "    | "    | "    | "    | "    | Amsterdam                | Proctitis                  |
| F/IC-Cal3  | 34 | 06   | 03   | 03   | 02   | 07   | 04   | 03   | California               | Cervicitis                 |
| F/8p       | 34 | "    | "    | "    | "    | "    | "    | "    | Lisbon                   | Cervicitis/urethritis      |
| F/9p       | 34 | "    | "    | "    | "    | "    | "    | "    | Lisbon                   | Cervicitis/urethritis      |
| E/19e      | 34 | "    | "    | "    | "    | "    | "    | "    | Ecuador                  | Cervicitis                 |
| E/5s       | 34 | "    | "    | "    | "    | "    | "    | "    | San Francisco            | Cervicitis/urethritis      |
| Ja/41nl    | 34 | "    | "    | "    | "    | "    | "    | "    | Amsterdam                | Cervicitis and vaginal discharge |
| Ja/47nl    | 34 | "    | "    | "    | "    | "    | "    | "    | Amsterdam                | Cervicitis and vaginal discharge |
| F/10s      | 35 | 06   | 03   | 03   | 02   | 07   | 04   | 08   | San Francisco            | PID                        |
| F/11s      | 35 | "    | "    | "    | "    | "    | "    | "    | San Francisco            | PID                        |
| F/12s      | 35 | "    | "    | "    | "    | "    | "    | "    | San Francisco            | PID                        |
| E/39nl     | 36 | 06   | 03   | 03   | 03   | 07   | 04   | 03   | Amsterdam                | Cervicitis                 |
| Da/TW448   | 37 | 06   | 03   | 03   | 07   | 04   | 02   | 02   | Taiwan                   | Trachoma                   |
| D/3e       | 38 | 06   | 03   | 06   | 02   | 07   | 04   | 03   | San Francisco            | Cervicitis/urethritis      |
| E/Bour     | 39 | 06   | 04   | 03   | 02   | 07   | 04   | 03   | California               | Cervicitis                 |
| E/45nl     | 39 | "    | "    | "    | "    | "    | "    | "    | Amsterdam                | Cervicitis                 |
| E/28e      | 39 | "    | "    | "    | "    | "    | "    | "    | Ecuador                  | Cervicitis                 |
| E/55e      | 39 | "    | "    | "    | "    | "    | "    | "    | Ecuador                  | Cervicitis                 |
| E/56e      | 39 | "    | "    | "    | "    | "    | "    | "    | Ecuador                  | Cervicitis                 |
| E/6p       | 39 | "    | "    | "    | "    | "    | "    | "    | Lisbon                   | Cervicitis                 |
| E/7p       | 39 | "    | "    | "    | "    | "    | "    | "    | Lisbon                   | Cervicitis                 |
| E/58t      | 39 | "    | "    | "    | "    | "    | "    | "    | Tanzania                 | Conjunctivitis             |
| E/4s       | 40 | 06   | 04   | 03   | 03   | 07   | 04   | 03   | San Francisco            | Cervicitis/urethritis      |
| Ja/26s     | 41 | 06   | 04   | 03   | 05   | 06   | 04   | 08   | San Francisco            | Cervicitis/urethritis      |
| C/31n      | 42 | 07   | 03   | 03   | 01   | 03   | 05   | 07   | Nepal                    | Trachoma, TI               |
| C/35n      | 43 | 07   | 03   | 03   | 06   | 03   | 05   | 07   | Nepal                    | Trachoma, TI               |
| C/1n       | 44 | 07   | 03   | 03   | 06   | 03   | 06   | 07   | Nepal                    | Trachoma, TI               |
| C/29n      | 44 | "    | "    | "    | "    | "    | "    | "    | Nepal                    | Trachoma, TI               |
| C/30n      | 44 | "    | "    | "    | "    | "    | "    | "    | Nepal                    | Trachoma, TI               |
| C/34n      | 44 | "    | "    | "    | "    | "    | "    | "    | Nepal                    | Trachoma, TI               |
| C/36n      | 44 | "    | "    | "    | "    | "    | "    | "    | Nepal                    | Trachoma, TI               |
| C/37n      | 44 | "    | "    | "    | "    | "    | "    | "    | Nepal                    | Trachoma, TI               |

*LGV, lymphogranuloma venereum; TS, trachomatous scarring; PID, pelvic inflammatory disease; TI, trachomatous inflammation severe; ST, sequence type.
†Strain ID, first letter refers to the ompA genotype; the number after the dash represents the ID# of the clinical strain; the small case letter after the number denotes the geographic region from which the sample was obtained: e, Ecuador; nl, the Netherlands; n, Nepal; p, Portugal; s, San Francisco.
eBURST report for 87 isolates by group*

**Group 1: No. isolates = 29 | No. STs = 19 | Predicted founder = 19**

| ST   | FREQ | SLV | DLV | TLV | SAT | Average distance | Group | Subgroup |
|------|------|-----|-----|-----|-----|------------------|-------|----------|
| ST19 | 5    | 9   | 9   | 0   | 0   | 1.5              | 87%   | 98%      |
| ST23 | 3    | 7   | 9   | 2   | 0   | 1.72             | 41%   | 87%      |
| ST21 | 2    | 5   | 9   | 4   | 0   | 1.94             | 14%   | 41%      |
| ST27 | 1    | 4   | 9   | 5   | 0   | 2.05             | 4%    | 19%      |
| ST29 | 1    | 4   | 6   | 7   | 1   | 2.27             | 3%    | 10%      |
| ST20 | 1    | 3   | 11  | 4   | 0   | 2.05             | 0%    | 0%       |
| ST24 | 2    | 3   | 9   | 6   | 0   | 2.16             | 0%    | 3%       |
| ST8  | 1    | 3   | 8   | 7   | 0   | 2.22             | 2%    | 0%       |
| ST5  | 1    | 3   | 6   | 8   | 1   | 2.38             | 0%    | 0%       |
| ST31 | 1    | 3   | 6   | 8   | 1   | 2.38             | 0%    | 0%       |
| ST16 | 1    | 3   | 5   | 10  | 0   | 2.38             | 0%    | 3%       |
| ST17 | 1    | 2   | 14  | 2   | 0   | 2.0              | 0%    | 0%       |
| ST30 | 1    | 2   | 11  | 5   | 0   | 2.16             | 0%    | 0%       |
| ST7  | 1    | 2   | 9   | 7   | 0   | 2.27             | 0%    | 0%       |
| ST9  | 2    | 2   | 7   | 8   | 1   | 2.44             | 0%    | 0%       |
| ST15 | 2    | 2   | 5   | 11  | 0   | 2.5              | 0%    | 0%       |
| ST3  | 1    | 1   | 8   | 9   | 0   | 2.44             | 0%    | 0%       |
| ST14 | 1    | 1   | 6   | 8   | 3   | 2.72             | 0%    | 0%       |
| ST29 | 1    | 1   | 5   | 11  | 1   | 2.66             | 0%    | 0%       |

*ST, sequence type; FREQ, frequency; SLV, single locus variant; DLV, double locus variant; TLV, triple locus variant; SAT, satellite

**Group 2: No. isolates = 19 | No. STs = 7 | Predicted founder = 11**

| ST   | FREQ | SLV | DLV | TLV | SAT | Average distance | ST Bootstrap |
|------|------|-----|-----|-----|-----|------------------|--------------|
| ST11 | 1    | 3   | 3   | 0   | 0   | 1.5              | 46%          | 17%         |
| ST43 | 1    | 3   | 2   | 1   | 0   | 1.66             | 37%          | 21%         |
| ST13 | 2    | 2   | 3   | 1   | 0   | 1.83             | 11%          | 0%          |
| ST12 | 7    | 2   | 2   | 2   | 0   | 2.0              | 9%           | 0%          |
| ST44 | 6    | 2   | 2   | 2   | 0   | 2.0              | 9%           | 0%          |
| ST42 | 1    | 1   | 2   | 2   | 1   | 2.5              | 0%           | 0%          |
| ST10 | 1    | 1   | 2   | 2   | 1   | 2.5              | 0%           | 0%          |

**Group 3: No. isolates = 22 | No. STs = 7 | Predicted founder = 34**

| ST   | FREQ | SLV | DLV | TLV | SAT | Average distance | ST Bootstrap |
|------|------|-----|-----|-----|-----|------------------|--------------|
| ST34 | 7    | 4   | 2   | 0   | 0   | 1.33             | 71%          | 55%         |
| ST39 | 8    | 3   | 3   | 0   | 0   | 1.5              | 34%          | 19%         |
| ST36 | 1    | 2   | 3   | 1   | 0   | 1.83             | 3%           | 0%          |
| ST40 | 1    | 2   | 2   | 2   | 0   | 2.0              | 8%           | 0%          |
| ST35 | 3    | 1   | 3   | 2   | 0   | 2.16             | 0%           | 0%          |
| ST38 | 1    | 1   | 3   | 2   | 0   | 2.16             | 0%           | 0%          |
| ST32 | 1    | 1   | 2   | 3   | 0   | 2.33             | 0%           | 0%          |

Singletons: size 11

- ST26
- ST25
- ST22
- ST18
- ST6
- ST4
- ST2
- ST1
- ST41
- ST37
- ST33

*No. of STs, 44; No. of resamplings for bootstrapping 1,000; No. of groups, 3.