Title

Chlamydia trachomatis Strain Types Have Diversified Regionally and Globally with Evidence for Recombination across Geographic Divides

Authors

Smelov, Vitaly
Vrbanac, Alison
van Ess, Eleanne F
et al.

Publication Date

2017

DOI

10.3389/fmicb.2017.02195

Peer reviewed

frontiers in Microbiology

ORIGINAL RESEARCH
published: 13 November 2017
doi: 10.3389/fmicb.2017.02195

Chlamydia trachomatis Strain Types Have Diversified Regionally and Globally with Evidence for Recombination across Geographic Divides

*Vitaly Smelov[1,2,3‡], Alison Vrbanac[4†‡], Eleanne F. van Ess[5‡], Marlies P. Noz[5], Raymond Wan[4], Carina Eklund[2], Tyler Morgan[4], Lydia A. Shrier[6], Blake Sanders[4], Joakim Dillner[2], Henry J. C. de Vries[7,8,9], Servaas A. Morre[5,10] and Deborah Dean[4,11,12,13]**

[1] International Agency for Research on Cancer, World Health Organization, Lyon, France, [2] Karolinska Institute, Stockholm, Sweden, [3] North-Western State Medical University named after I.I. Mechnikov, St. Petersburg, Russia, [4] UCSF Benioff Children's Hospital Oakland Research Institute, Oakland, CA, United States, [5] Laboratory of Immunogenetics, Department of Medical Microbiology and Infection Control, VU University Medical Center, Amsterdam, Netherlands, [6] Department of Pediatrics, Boston Children's Hospital, Boston, MA, United States, [7] Center for Infection and Immunology Amsterdam, Academic Medical Center, Amsterdam, Netherlands, [8] Department of Dermatology, Academic Medical Center, Amsterdam, Netherlands, [9] STI Outpatient Clinic, Public Health Service of Amsterdam (GGD Amsterdam), Amsterdam, Netherlands, [10] Institute of Public Health Genomics, Department of Genetics and Cell Biology, Research Institute GROW (School for Oncology and Developmental Biology), Faculty of Health, Medicine and Life Sciences, University of Maastricht, Maastricht, Netherlands, [11] Department of Bioengineering, University of California, Berkeley, Berkeley, CA, United States, [12] Department of Bioengineering, University of California, San Francisco, San Francisco, CA, United States, [13] Departments of Medicine and Pediatrics, University of California, San Francisco, San Francisco, CA, United States

Chlamydia trachomatis (*Ct*) is the leading cause of bacterial sexually transmitted diseases worldwide. The *Ct* Multi Locus Sequence Typing (MLST) scheme is effective in differentiating strain types (ST), deciphering transmission patterns and treatment failure, and identifying recombinant strains. Here, we analyzed 323 reference and clinical samples, including 58 samples from Russia, an area that has not previously been represented in *Ct* typing schemes, to expand our knowledge of the global diversification of *Ct* STs. The 323 samples resolved into 84 unique STs, a 3.23 higher typing resolution compared to the gold standard single locus *omp*A genotyping. Our MLST scheme showed a high discriminatory index, *D,* of 0.98 (95% CI 0.97–0.99) confirming the validity of this method for typing. Phylogenetic analyses revealed distinct branches for the phenotypic diseases of lymphogranuloma venereum, urethritis and cervicitis, and a sub-branch for ocular trachoma. Consistent with these findings, single nucleotide polymorphisms were identified that significantly correlated with each phenotype. While the overall number of unique STs per region was comparable across geographies, the number of STs was greater for Russia with a significantly higher ST/sample ratio of 0.45 (95% CI: 0.35–0.53) compared to Europe or the Americas ($p < 0.009$), which may reflect a higher level of sexual mixing with the introduction of STs from other regions and/or reassortment of alleles. Four STs were found to be significantly associated with a particular geographic region. ST23 [$p = 0.032$ (95% CI: 1–23)], ST34 [$p = 0.019$ (95% CI: 1.1–25)]; and ST19 [$p = 0.001$ (95% CI: 1.7–34.7)] were significantly associated

OPEN ACCESS

Edited by:
Jorge Blanco, Universidade de Santiago de Compostela, Spain

Reviewed by:
Wilhelmina May Huston, University of Technology Sydney, Australia
Fernando Gonzalez-Candelas, Universitat de València, Spain
Carolyn Black, National Center for Emerging and Zoonotic Infectious Diseases (CDC), United States

***Correspondence:**
Deborah Dean
ddean@chori.org

†Present address:
Alison Vrbanac, Center for Immunity, Infection and Inflammation, University of California, San Diego, San Diego, CA, United States

‡These authors have contributed equally to this work.

Specialty section:
This article was submitted to Infectious Diseases, a section of the journal Frontiers in Microbiology

Received: *27 July 2017*
Accepted: *25 October 2017*
Published: *13 November 2017*

Citation:
Smelov V, Vrbanac A, van Ess EF, Noz MP, Wan R, Eklund C, Morgan T, Shrier LA, Sanders B, Dillner J, de Vries HJC, Morre SA and Dean D (2017) Chlamydia trachomatis Strain Types Have Diversified Regionally and Globally with Evidence for Recombination across Geographic Divides. Front. Microbiol. 8:2195. doi: 10.3389/fmicb.2017.02195

with Netherlands compared to Russia or the Americas, while ST 30 [$p = 0.031$ (95% CI: 1.1–17.8)] was significantly associated with the Americas. ST19 was significantly associated with Netherlands and Russia compared with the Americans [$p = 0.001$ (95% CI: 1.7–34.7) and $p = 0.006$ (95% CI: 1.5–34.6), respectively]. Additionally, recombinant strains were ubiquitous in the data set [106 (32.8%)], although Europe had a significantly higher number than Russia or the Americas ($p < 0.04$), the majority of which were from Amsterdam [43 (87.8%) of 49)]. The higher number of recombinants in Europe indicates selective pressure and/or adaptive diversification that will require additional studies to elucidate.

Keywords: *Chlamydia trachomatis*, MLST, recombination, global diversification, allele mixing, reassortment

INTRODUCTION

The modern pathogenic *Chlamydiaceae* family has a rich evolutionary history, diverging from environmental *Chlamydiales* approximately seven million years ago (Horn et al., 2004). The human *Chlamydiaceae* spp. *Chlamydia trachomatis* (*Ct*) has infected human populations causing sexually transmitted diseases (STD) and the chronic ocular disease known as trachoma since the 27th century BC (Perkins and Hill, 2013). Trachoma was initially described in China and in the Eber's Papyrus of Egypt, and subsequently spread to Europe during the Crusades (Perkins and Hill, 2013). While improvements in hygiene and sanitation have eliminated trachoma from many global populations, the disease is still endemic in many developing countries of Africa, Central and South America, the South Pacific and Asia in addition to aboriginal populations in Australia (Perkins and Hill, 2013). Presently, *Ct* is the leading cause of preventable blindness and bacterial sexually transmitted infections (STIs) worldwide (Rowley et al., 2012) with estimates of over 250 million trachoma cases and 110 million annual STI cases, according to the World Health Organization (WHO Sexually transmitted infections [STIs], 2015).

Ct has evolved to include 19 serological variants (serovars) based on antibody typing of the major outer membrane protein (MOMP) with over 60 *omp*A genotypes (Dean and Millman, 1997; Batteiger et al., 2014; Isaksson et al., 2016; Peuchant et al., 2016; Schillinger et al., 2016), the gold standard typing technique for all *Chlamydiaceae* spp. The serovars are designated A through K, Ba, Da, Ia, Ja, and L_{1-3}, and L_2a while the *omp*A genotypes or strains are denoted by the same or by a number or letter after the conventional serovar name (e.g., D1; Ga) for new genotypes (Batteiger et al., 2014). These strains are responsible for ocular, urogenital and rectal infections. Ocular infections include trachoma, a chronic ocular disease, and ophthalmia neonatorum (Darville, 2005), an infection acquired during passage through a *Ct* infected birth canal (WHO Sexually transmitted infections [STIs], 2015). Urogenital strains cause not only ocular infections, which usually present as unilateral conjunctivitis (Dean et al., 2008), but also can ascend from the endocervix to cause sequelae such as pelvic inflammatory disease, infertility and ectopic pregnancy (Mårdh, 2004; Blas et al., 2007; Baud et al., 2008). Rectal infections can progress to proctitis and inguinal syndrome (Sethi et al., 2009). While the later is caused primarily by the lymphogranuloma venereum strains (LGV) L_{1-3}, L_2a, L_2b, and L_2c, the former can be caused by most *Ct* strains, although strains B, Ba, and C are rarely detected in the urethra, endocervix or rectum (Danby et al., 2016; Labiran et al., 2016). Strain A is the only strain that is confined to the ocular mucosa (Dean, 2010).

Classification of *Ct* strains was conventionally performed by serotyping and more recently by *omp*A genotyping (Dean et al., 1992) since the organism is rarely cultured, a requirement for serotyping. Although *omp*A genotyping can be informative, the gene represents a mere 0.1% of the genome and is subject to immune selective pressure and recombination (Joseph et al., 2011, 2012). Finer, more holistic typing schemes are necessary to track recombination events, differentiate new and persistent infections (Götz et al., 2013), reinfection, transmission patterns and elucidate potential biomarkers. Three Multi Locus Sequence Typing (MLST) schemes have been developed for *Ct* (Klint et al., 2007; Pannekoek et al., 2008; Dean et al., 2009), of which only two meet the MLST criteria of using strictly housekeeping genes (Pannekoek et al., 2008; Dean et al., 2009). Our MLST scheme employs seven highly conserved housekeeping genes and successfully resolves reference and clinical *Ct* samples into LGV, trachoma, non-prevalent non-invasive urogenital, and prevalent non-invasive urogenital clonal complexes representing the respective diseases in addition to revealing evidence for recombination (Dean et al., 2009; Batteiger et al., 2014).

Partial and whole genome sequencing (WGS) have added considerably to our knowledge of the diversity of *Ct* and evidence for recombination. We initially bioinformatically identified recombination within *omp*A (Millman et al., 2001) and then among partial genome sequences of trachoma and sexually transmitted strains involving *omp*A and polymorphic membrane proteins (*pmp*) (Gomes et al., 2004, 2006, 2007). In the first publication of comparative WGS of *Ct*, we identified three major clades that were similar to the MLST disease-related clonal complexes with a subclade encompassing the trachoma strains within the non-prevalent non-invasive urogenital clade (Joseph et al., 2011). Subsequent WGS by our and other groups have substantiated these phylogenetic groupings as well as the recombinogenic nature of *Ct* (Harris et al., 2012; Joseph et al., 2012; Seth-Smith et al., 2013; Hadfield et al., 2017).

As whole genome sequencing remains cost-prohibitive for large sample sets and beyond the reach of most research

laboratories, in this work, 323 *Ct* reference and clinical samples from 15 countries and 5 continents were analyzed by MLST to provide a more comprehensive analysis of the global diversification of *Ct* strain types. We included 60 new clinical samples from Amsterdam, Netherlands, eight from Boston, MA, United States, and 58 from St. Petersburg, Russia, a region that had not previously been evaluated by MLST.

MATERIALS AND METHODS

Study Populations and Ethics

Information on populations and ethics for samples collected previously are included in the publications by Dean et al. (2009) and Batteiger et al. (2014). Russian women were enrolled in the original studies following verbal informed consent after approval by the Local Institutional Review Board at DO Ott Institute of Obstetrics and Gynecology and the Russian Academy of Medical Sciences (RAMS), St. Petersburg, Russia (Shipitsyna et al., 2007). The studies were additionally approved by the Department of Clinical Investigations and Intellectual Property, St. Petersburg Medical Academy of Postgraduate Studies (North-Western State Medical University named after I.I. Mechnikov since 2011), under the Federal Agency of Public Health and Social Development of Roszdrav, in accordance with the Declaration of Helsinki. For the 60 Dutch samples, each was obtained as previously described (Dean et al., 2009) according to the Declaration of Helsinki. The eight Boston samples were sent to Dr. Dean as de-identified samples with no trace to patient name and were not considered human subjects according to IRB at Children's Hospital Oakland Research Institute and NIH guidelines.

For the Russian samples, endocervical swabs were obtained consecutively from January 2006 to January 2008 in two university clinics in St. Petersburg, Russia, as described elsewhere (Smelov et al., 2009). To minimize the potential for low response from the enrolled women, samples were collected without obtaining personal information. After removal of any mucopus with a cotton swab, a Dacron swab was inserted into the endocervix, rotated and placed in an empty 5 mL vials. Specimens were kept at 4°C (39°F) for up to 4 days before they were shipped to the laboratory where they were stored at 4°C (39°F). Within 1–3 days (Shipitsyna et al., 2007) all samples were tested for the presence of *Ct* by commercial NAAT assays (Shalepo et al., 2006; Smelov et al., 2009). Additionally, either a conventional PCR (Lytech, Moscow, Russia) or a real-time PCR (Central Research Institute of Epidemiology, Moscow, Russia) were used. The results were confirmed in Amsterdam, Netherlands by the commercial real-time PCR (TaqMan, Applied Biosystems, United States) (Morré et al., 1999; Smelov et al., 2009) and CE-IVD certificated Presto CT-NG Assays (Goffin Molecular Technologies, Beek, Netherlands) (de Waaij et al., 2015).

Ct Reference and Clinical Samples

A total of 323 reference and clinical samples were analyzed that included 58 endocervical samples from St. Petersburg, 60 additional endocervical samples from Amsterdam and eight endocervical samples from Boston, Massachusetts; 265 were already in the MLST database[1] including the 20 reference strains: A/SA-1, A/HAR13, B/TW5/OT, Ba/Apache2, C/TW3/OT, D/UW3/Cx, Da/TW-448, E/Bour, F/IC-Cal3, G/UW57/Cx, H/UW4/Cx, I/UW12/Ur, Ia/UW202, J/UW36/Cx, Ja/UW92, K/UW36/Cx, L_1/440, L_2/434, L_{2a}/UW396, L_3/404.

ompA Genotyping and MLST Analysis

Genomic DNA was purified from clinical isolates and *ompA* genotyped as described previously (Dean et al., 2009; Batteiger et al., 2014). Only the samples from St. Petersburg were provided as swabs; DNA was extracted and purified for these samples using our protocol as described in Joseph et al. (2014). MLST analysis examined seven housekeeping genes: *glyA, mdhC, pdhA, yhbG, pykF, lysS,* and *leuS*, with primers as described (Dean et al., 2009) (Supplementary Table 1). All seven housekeeping genes were amplified and sequenced as described (Batteiger et al., 2014). A consensus sequence was created from the forward and reverse sequence. The genes for each of the St. Petersburg, Amsterdam and Boston samples were each concatenated and queried against the 265 samples in the MLST database in addition to including these samples in the database. Allelic numbers and STs were assigned based on this query as described previously (Batteiger et al., 2014).

Phylogenetic Analysis and Strain Clustering

Using the concatenated sequences, dataset strain clustering and Single Nucleotide Polymorphism (SNP) analyses were performed as described (Batteiger et al., 2014). Briefly, this included visualizing clusters of related STs and non-related STs using eBURST[2] (Feil et al., 2004). Founder STs were identified by the highest number of single locus variants (SLV) branching from that particular ST (i.e., the clonal ancestor that diversifies into other STs). Clonal complexes generated by eBURST were defined as a group of STs separated by one SLV.

Phylogenetic trees were created by Maximum likelihood using the Symmetric+GI model, which provided the best fit for the data, in the R package phangorn (Schliep, 2010) to analyze the nucleotide sequence variation between the seven MLST loci for each ST. Tree nodes were verified with 1,000 bootstrap replicates. Alternative evolutionary pathways, such as horizontal gene transfer, were analyzed with SplitsTree[3] using the splits decomposition method as described (Dean et al., 2009; Tamura et al., 2013). In addition, the sequence for each of the seven MLST loci for a sample were compared across the dataset and to the *ompA* genotype of the same sample to determine evidence for putative recombination.

Statistical Tests

Fisher's exact test was performed in R[4] to test for significant region-specific *ompA* and ST clustering; a p-value of <0.05 was

[1] https://pubmlst.org/chlamydiales/

[2] http://eburst.mlst.net

[3] www.splitstree.org

[4] http://www.r-project.org

considered significant. Confidence intervals were determined based on the method of Clopper-Pearson (Borkowf, 2006). Simpson's Diversity Index, *D*, was calculated for the MLST data as described (Simpson, 1949; Hunter and Gaston, 1988). A *D*-value of ≥0.95 was considered ideal for molecular typing techniques (van Belkum et al., 2007). The *omp*A genotypes were excluded from the analysis. The Benjamin-Hochberg FDR method (Benjamini and Hochberg, 1995) was used to correct *p*-values for multiple comparisons.

Samples were classified as putative recombinants when the sequences of the seven gene sequences that comprise the ST or any of the seven genes were non-concordant with each other or with the *omp*A genotype of the same sample.

PROC FREQ in SAS was used to identify SNPs associated with disease phenotype and Haplotype as described previously (Dean et al., 2009). Levene's test evaluated the variance across the dataset of the 323 samples. The Classification Index was used to determine significance of each SNP with a disease phenotype where a *p*-value of <0.05 was considered significant.

DnaSP v5.10 (Librado and Rozas, 2009) was used to calculate nucleotide (nt) and haplotype (hd) to determine the genetic diversity and differentiation for regional subgroups on the concatenated sequences of the seven MLST genes. DnaSP considers the frequency of variants (STs) present in a population and also genetic distances that separate these variants from each other. Genetic population differentiation between regional subgroups was assessed using the pairwise fixation index (Fst) in Arlequin v3.5 (Excoffier and Lischer, 2010) with significance testing by permutation.

RESULTS

Characteristics and Geographic Distribution of Alleles

The characteristics of the alleles for each gene locus are shown in Supplementary Table 2. The number of alleles varied by gene locus, ranging from seven to 18, as did the number of polymorphic sites. We determined the allele frequencies by geographic region for the 78 alleles (**Table 1**). Thirty-two (41%) alleles were observed once. The highest number of unique alleles for a geographic region was in Western Europe at 16 alleles but the highest frequency was 43.3% for Russia, which was not statistically significant.

ST and *omp*A Distributions

For the 323 samples, 84 unique STs were identified (Supplementary Table 3). STs novel to the dataset were numbered consecutively in order of identification. Of those 84, 57 (67.9%) were singletons, with a relatively even distribution by geographic region (excluding Asia and Africa where the sample sizes were very small) with a higher percentage of singletons in Europe that was not significantly different (**Table 2**; $p = 0.08$). **Table 2** also shows that the percentage of unique STs per region was highly similar. However, the ST/sample ratio was significantly greater for Russian than for European and American samples ($p < 0.009$). There were also significant differences in the distribution of STs (**Table 3**). Dutch females ($n = 79$) were significantly more likely to be infected with ST23 ($p = 0.032$; 95% CI: 1–23) and ST34 ($p = 0.02$; 95% CI: 1.1–25) compared to Russian females ($n = 58$) and with ST19 ($p = 0.001$; 95% CI: 1.7–34.7) compared to American females ($n = 108$). Supplementary Table 4 shows the distribution of STs by geographic region.

There were 26 *omp*A genotypes observed in the dataset resulting in a 3.23 lower resolution than for STs. Excluding the STDs samples from South Africa, all samples available from Asia and Africa were from trachoma patients; 1 (7.1%) of 14 was a urogenital Da *omp*A genotype (ST 37) in Asia while 1 (5.9%) of 17 was a urogenital E genotype (ST 39) in Africa.

*omp*A genotype E ($n = 72$) was the most prevalent and associated with 18 STs with an ST to sample ratio of 0.25 (Supplementary Table 3). The distribution for the remaining *omp*A genotypes in descending order was G (38; 16 STs; ratio: 0.42); D (33; 16 STs; ratio: 0.48); F (30; 5 STs; ratio: 0.17); Ia (17; 4 STs; ratio: 0.25); K (15; 7 STs; ratio: 0.47); J (13; 9 STs; ratio: 0.69); B, H and I at 12 samples each (7, 6, and 3 STs, respectively; ratios: 0.58, 0.50, and.25, respectively); C (11: 5 STs; ratio: 0.45); A (9; 4 STs; ratio: 0.44); D2 and Ja at 4 samples each (1 and 3 STs, respectively; ratios: 0.25 and 0.75, respectively); Ba, D1, E6, F4 and Ia4 at 2 samples each (1 ST for each; ratio: 0.50 for each); and Da (1; 1 ST; ratio: 0.10).

The *omp*A distribution of urogenital strains varied across Europe, the Americas, and Russia. In Europe, *omp*A genotype D was significantly more prevalent than in other geographic regions ($p = 0.046$) and more frequent than the globally prevalent genotype E. In Russia, *omp*A genotype G was significantly more prevalent than in all other regions ($p = 0.001$). In the Americas, *omp*A genotype Ia was significantly more prevalent than the other regions ($p = 0.001$). In comparing female STD cohorts, Russian women were significantly more likely to be infected with E and G ($p = 0.001$ and 0.026, respectively) while Dutch women were significantly more likely to be infected with D and I ($p = 0.026$ and 0.002, respectively).

Simpson's Discriminatory Index was 0.98 (95% CI 0.97–0.99) for MLST and 0.70 (95% CI 0.67–0.73) for *omp*A genotyping.

Based on DnaSP, nucleotide diversity was lower in Russia (Pi = 0.00216) compared to Netherlands (Pi = 0.00321) and North America (Pi = 0.00294) (**Table 4**). However, Netherlands and North American datasets contained isolates from men and LGV samples. The phylogenetically disparate LGV samples appeared to contribute substantially to nucleotide diversity; nucleotide diversity dropped from 0.00250 to 0.00196 in Dutch women when the four LGV isolates were removed. When comparing non-LGV isolates from women in these regions, Russian women had the highest nucleotide diversity (Pi = 0.00216), followed by North American women (Pi = 0.00210), and Dutch women (Pi = 0.00196).

Assessing population differentiation between regional subgroups by Fst revealed significant differences between most regional subgroup pairwise comparisons (**Table 5**). African and Asian subgroups exhibited high Fst values in all comparisons, indicating those regions were distinct from others in the dataset. This is not surprising given the small sample size and the fact that these samples are from two distinct trachoma populations.

TABLE 1 | Allele frequencies by geographic region by locus.

Gene locus	No. alleles	Africa *n* = 21 (%)	Western Europe *n* = 109 (%)	Russia *n* = 58 (%)	Asia *n* = 13 (%)	Americas *n* = 122 (%)	Classification index *p*-value
glyA	12	1 (14.29)	1 (12.84)			1 (7.37)	<0.001
			3 (50.46)	3 (62.07)	3 (30.77)	**2 (0.82)**	
		3 (80.95)	**4 (0.92)**	6 (34.48)		3 (42.62)	
		6 (4.76)	**5 (1.83)**	**10 (1.72)**	6 (7.69)		
			6 (32.11)	**11 (1.72)**	**7 (61.54)**	6 (48.36)	
			8 (0.92)				
			9 (0.92)			**12 (0.82)**	
mdhC	7	1 (14.29)	1 (11.93)			1 (8.19)	<0.001
		3 (76.19)	**2 (2.75)**	3 (63.79)	3 (100.00)	3 (68.85)	
		4 (4.76)	3 (70.64)	4 (34.48)			
		5 (4.76)	4 (12.84)	**7 (1.72)**		4 (22.95)	
			6 (1.83)				
pdhA	8	**1 (4.76)**					<0.001
		3 (95.24)	**2 (0.92)**	3 (94.83)	3 (100)	3 (97.54)	
			3 (91.74)	5 (3.45)			
			4 (2.75)	**8 (1.72)**		5 (1.64)	
			5 (3.67)			**6 (0.82)**	
			7 (0.92)				
ybhG	11				**1 (7.69)**		<0.001
		2 (4.76)	2 (31.19)	2 (44.83)	**4 (7.69)**	2 (43.44)	
		6 (80.96)	3 (0.92)	6 (48.28)	5 (7.69)		
		8 (14.29)	5 (1.83)	9 (5.17)	6 (76.92)	3 (0.82)	
			6 (48.62)	**11 (1.72)**		5 (9.83)	
			7 (0.92)			6 (37.70)	
			8 (15.60)				
			9 (0.92)			8 (7.38)	
						10 (0.82)	
pykF	10	1 (14.29)	1 (15.60)			1 (8.20)	<0.001
		3 (71.43)	6 (49.54)	6 (53.45)	3 (92.31)	**2 (2.46)**	
		6 (9.52)	7 (33.03)	7 (37.93)		**4 (0.82)**	
		7 (4.76)	**8 (0.92)**	**10 (8.62)**	7 (7.69)	**5 (0.82)**	
			9 (0.92)			6 (36.07)	
						7 (51.64)	
lysS	12		1 (6.42)	1 (8.62)		1 (7.38)	<0.001
		4 (23.81)	4 (74.31)	4 (74.14)	4 (7.69)	**2 (0.82)**	
		5 (71.43)	8 (18.35)	8 (13.79)	5 (30.77)	**3 (0.82)**	
		7 (4.76)	**10 (0.92)**	**11 (1.72)**		4 (73.77)	
				12 (1.72)	**6 (61.54)**	8 (14.75)	
						9 (2.46)	
leusS	18					**1 (2.46)**	<0.001
		2 (4.76)	3 (75.23)	3 (79.31)	2 (7.69)	3 (67.21)	
		3 (4.76)	**4 (0.92)**	6 (8.62)	7 (84.62)		
		7 (9.52)	**5 (0.92)**	8 (3.45)		6 (5.74)	
		9 (66.67)	6 (3.67)	**14 (1.72)**	**10 (7.69)**	8 (14.75)	
		11 (14.29)	8 (2.75)	**15 (1.72)**			
			11 (15.60)	**16 (1.72)**		9 (0.82)	
			13 (0.92)	**17 (1.72)**		**11 (7.38)**	
				18 (1.72)		**12 (1.64)**	
No. novel alleles per region		3 (12.5%)	16 (40%)	13 (43.3%)	5 (29.4)	13 (36.1%)	
Total no. alleles	78	24	40	30	17	36	

**Numbers are arranged vertically for each locus to represent the individual alleles (e.g., glyA alleles are assigned 1, 2, 3, 4, 5, 6, 7…12 because there are 12 alleles for this locus). Alleles marked in boldface are specific for a single geographic region. n, number of samples.*

TABLE 2 | Strain type (ST) diversity and recombinants by geographic region.

Geographic location	Number of Samples	Number of ST's	ST/sample ratio (95% CI)	Number ST Singletons (%)	Number Novel STs per Region (%)	Number Recombinant Samples (%)
Europe	109	28	0.26 (0.18–0.39)	17 (61)	19 (68)	49 (45)**
Americas	122	34	0.28 (0.20–0.40)	16 (47)	24 (71)	28 (23)
Asia	13	8	0.57 (0.48–0.66)	5 (63)	8 (100)	1 (7.7)
Africa	21	7	0.35 (0.26–0.44)	5 (71)	5 (71)	8 (38.1)
Russia	58	26	0.45*(0.35–0.53)	14 (54)	18 (69)	20 (34.5)
Total	**323**			**57**		**106 (32.8)**

**Significantly increased compared to Europe and the Americas ($p < 0.009$); Asia and Africa were excluded due to small numbers. **Significantly increased compared to Russia and the Americas ($p < 0.04$).*

TABLE 3 | Differences in ST distribution by geographic region.

Regions of comparison*	ST	Significance
Netherlands vs. Russia	23	$p = 0.03199$ (95% CI: 1–23)
Netherlands vs. Russia	34	$p = 0.0188$ (95% CI: 1.1–25)
Netherlands vs. Americas	19	$p = 0.000399$ (95% CI: 1.7–34.7)
Americas vs. Netherlands	39	$p = 0.03127$ (95% CI: 1.1–17.8)
Russia vs. Americas	19	$p = 0.005752$ (95% CI: 1.5–34.6)

**Women were analyzed separately for the purposes of comparing similar cohorts.*

Though Western Europe was significantly different from all subgroups, the Russia and Americas subgroups were not significantly different by Fst. However, this could be attributed to the higher LGV representation in the Western European samples (15.5%) in comparison to the American samples (7.3%), and Russian samples (0%). Concordantly, when LGV samples were removed from the Western European dataset, Western Europe was no longer significantly different from Russia (**Table 5**). When evaluating only women without LGV from the Netherlands, Russia, and North America, the Russian women were not significantly different from the American and Netherlands women (Supplementary Table 8).

Phylogenetic Relationships and Evidence for Recombination

The phylogenetic relationships were initially evaluated by eBURST, which revealed clonal clusters (CC) similar to what we reported previously (**Figure 1**) (Dean et al., 2009; Batteiger et al., 2014) but with the addition of an LGV cluster. These included CC-A encompassing trachoma STs, CC-B with non-invasive, non-prevalent urogenital STs, CC-C with non-invasive prevalent STs and CC-D that included LGV STs. The predicted founders were ST19, ST23, ST34, and ST39 (**Figure 1** and Supplementary Table 5). The 57 singleton STs are denoted by a gray circle alone or within a colored circle, representing a specific geographic region, but were not associated with any specific region.

The tree (**Figure 2**) revealed ST branches similar to the eBURST clusters; both had branches or clusters for the disease phenotypic groups of LGV, non-invasive prevalent urogenital, and non-invasive non-prevalent urogenital STs. However, the trachoma STs formed a subgroup of the non-invasive non-prevalent urogenital branch. In addition, within disease groups, STs branched from central nodes by geographic region. These nodes contained, in general, large numbers of STs from diverse locations. For example, the founders ST19, ST23, ST34, and ST39 located at nodes in the tree contain STs from Europe, Russia, and the Americas. The amino acid tree showed a similar phylogeny (Supplementary Figure 1).

The Splitstree decomposition tree revealed evidence for a network structure consistent with homologous recombination (**Figure 3**). This was demonstrated by the interconnecting networks specifically among the ST founders ST19, ST23, ST34, and ST39 and other STs on the network in addition to the canonical evolutionary pathway shown in the tree (**Figure 2**). Supplementary Figure 2 shows the amino acid tree.

The Splitstree data are consistent with findings in the MLST and *omp*A genotyping data. Samples were classified as recombinant when the sequences of the seven genes that determine the ST were non-concordant with each other or with the seven gene sequences of the *Ct* genotype associated with the *omp*A genotype (**Table 6**), denoted in bold in Supplementary Table 3. There was no evidence of any recombination among the seven ST genes for any sample, although this would conceivably be possible.

A total of 106 (32.8%) samples were considered putative recombinants. Excluding Asia and Africa where the sample sizes were small, Europe had a significantly higher number of recombinants than Russia and the Americas ($p < 0.04$) (**Table 2**), where, of the 109 samples, 49 were recombinant with 43 (87.8%) from the Netherlands. While *omp*A genotypes were generally consistent across samples of the same ST, many cases of recombinant strains were observed. For example, ST19 ($n = 40$) was primarily associated with *omp*A genotype G (37.5%) but eight different *omp*A genotypes (B, D, E, G, H, I, J, K) were also associated with this ST. The majority of these samples were from St. Petersburg and Amsterdam (Supplementary Table 3). ST23 ($n = 32$) was also associated with eight *omp*A genotypes (B, D, G, H, I, Ia, J, K) but the most frequent was Ia (47%). In contrast, the most geographically prevalent ST was ST39 ($n = 45$) where 95% were associated with *omp*A genotypes E. There were also 24 singletons that were recombinants (**Table 6**).

Table 7 shows the *omp*A genotype, ST and allelic SNPs, if present, associated with each of the eight Boston samples added to the dataset. Under the column denoted ST sequence homology are the *Ct omp*A genotype sequences to which the sequences of the seven MLST genes are identical. For example, sample J/259b

TABLE 4 | Nucleotide and haplotype diversity.

Subgroup	*N*	Nucleotide diversity(nt)	Haplotype diversity(hd)
Russia	58	0.00216 ± 0.00007	0.917 ± 0.023
Netherlands	79	0.00321 ± 0.00030	0.879 ± 0.020
North America	116	0.00294 ± 0.00023	0.926 ± 0.012
North American Women	66	0.00210 ± 0.00009	0.925 ± 0.016
Netherlands Women	72	0.00250 ± 0.00026	0.863 ± 0.024
Netherlands Women non-LGV	68	0.00196 ± 0.00011	0.846 ± 0.026

TABLE 5 | Pairwise population differentiation (Fst) for regional subgroups.

	Africa	Asia	Americas	Russia	Western Europe	Western Europe non-LGV	Americas non-LGV
Africa		0.23421*	0.27401*	0.33802*	0.17265*		
Asia			0.36715*	0.44728*	0.29367*		
Americas				0.01482 ns	0.02892*		
Russia					0.05425*	0.00305 ns	0.00884 ns
Western Europe							
Western Europe non-LGV							0.02627*
Americas non-LGV							

*$p < 0.05$ by permutation test ($n = 110$).

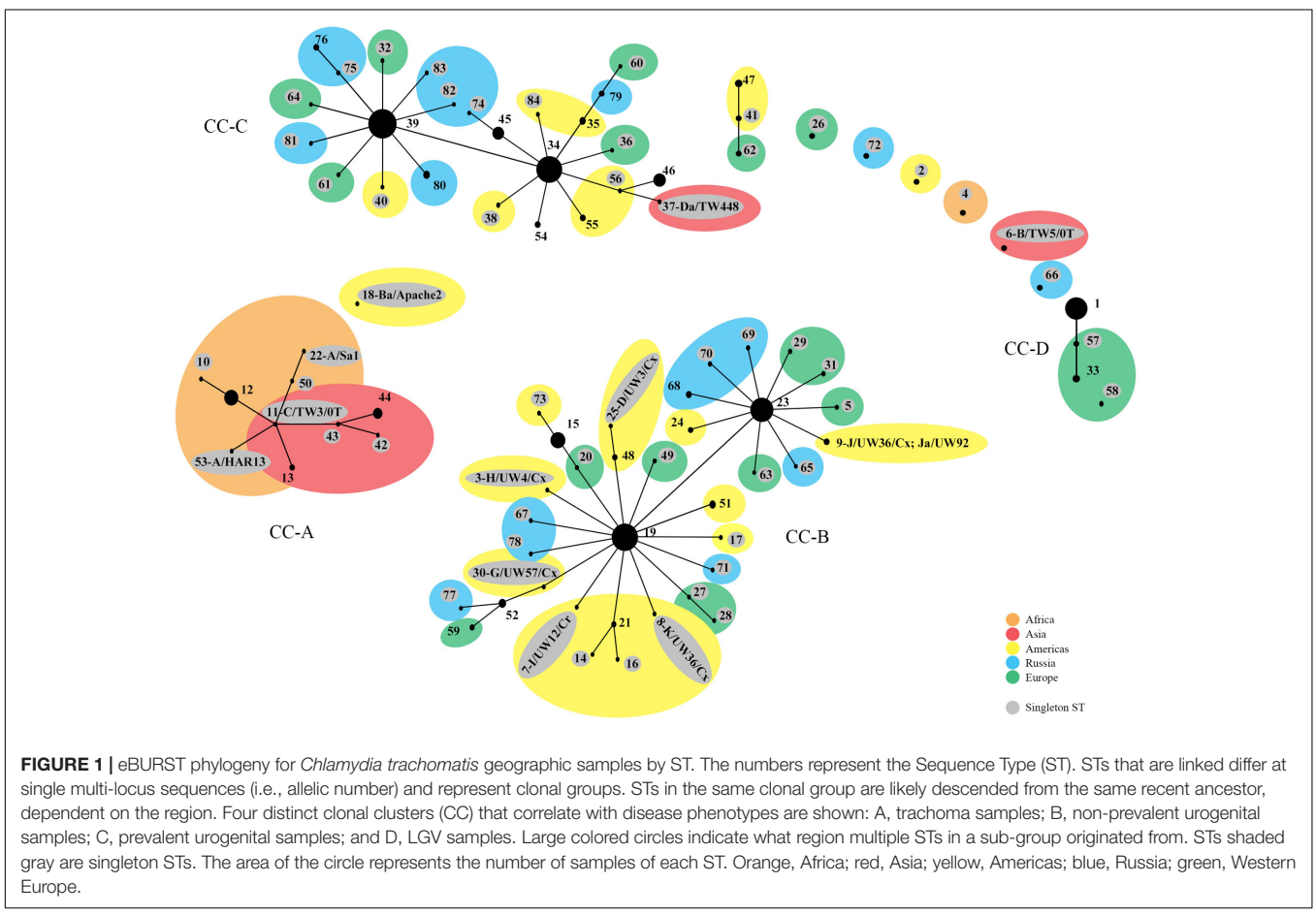


FIGURE 1 | eBURST phylogeny for *Chlamydia trachomatis* geographic samples by ST. The numbers represent the Sequence Type (ST). STs that are linked differ at single multi-locus sequences (i.e., allelic number) and represent clonal groups. STs in the same clonal group are likely descended from the same recent ancestor, dependent on the region. Four distinct clonal clusters (CC) that correlate with disease phenotypes are shown: A, trachoma samples; B, non-prevalent urogenital samples; C, prevalent urogenital samples; and D, LGV samples. Large colored circles indicate what region multiple STs in a sub-group originated from. STs shaded gray are singleton STs. The area of the circle represents the number of samples of each ST. Orange, Africa; red, Asia; yellow, Americas; blue, Russia; green, Western Europe.

has four SNPs that are different from the sequences of the seven genes for reference and clinical J strains in the database. These SNPs are identical to the sequences of strains G and K, which suggests that this sample is a recombinant between J and G or K strains. Another example is sample D/256b; the *omp*A genotype is D but the MLST sequences of the seven genes are identical to

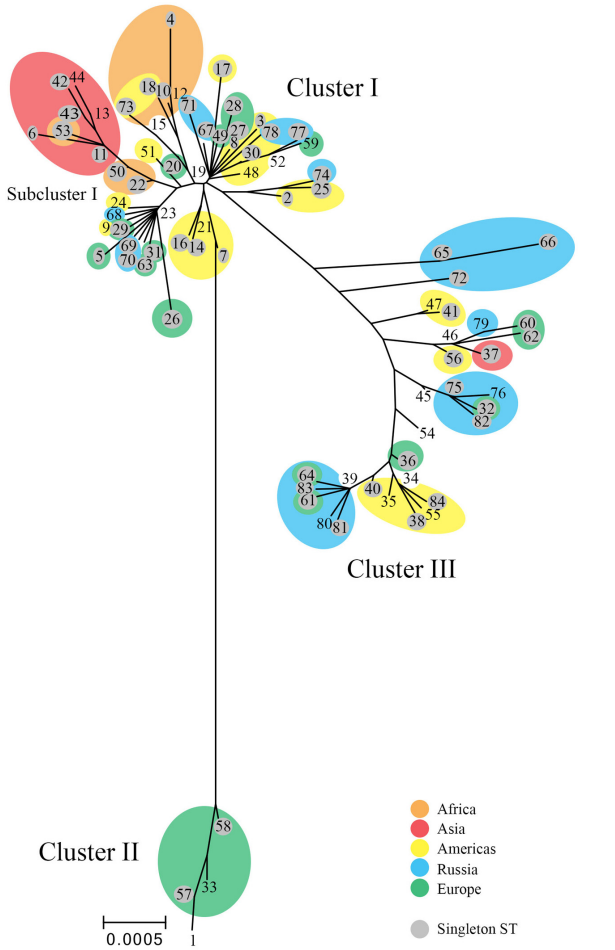


FIGURE 2 | Phylogenetic tree for *C. trachomatis* nucleotide sequences by ST. The numbers represent the ST. The tree was constructed using the matrix of pairwise differences between the 84 ST concatenated nucleotide sequences for the seven MLST loci using the maximum composite likelihood method for estimating genetic distances. Only bootstrap values (1,000 replicates) of >70% are shown at branch points in the tree. STs (shaded gray) are singletons. Orange, Africa; red, Asia; yellow, Americas; blue, Russia; green, Western Europe. Scale bar indicates number of substitutions per site.

reference and clinical F strains in the database. Therefore, it is a recombinant between D and F. Five (62.5%) of the eight Boston samples were putative recombinants. Supplementary Table 5 shows similar data for the 60 new samples from Amsterdam. For the Amsterdam samples, 43 (54.4%) of 79 samples were recombinants. Likewise, Supplementary Table 6 represents the data from St. Petersburg where 20 (34.5%) of 58 samples were recombinants.

Disease Phenotypes Are Associated with Specific Single Nucleotide Polymorphisms (SNPs)

Table 8 shows the SNPs that are specific for each Haplotype and disease phenotype group. These groups include invasive urogenital disease caused by LGV strains (Haplotype 1), non-invasive urogenital disease group that includes the most prevalent worldwide strains including D, Da, E, and F (Haplotype 2), and trachoma caused by ocular A, B, Ba, and C strains (Haplotype 3). The non-invasive group includes clinical Ja strains that are recombinants of E and F strains. There are 14 SNPs that are specific for the LGV disease group, and each SNP independently identifies this group and Haplotype. The same is true for the non-invasive disease group. However, the two SNPs for the trachoma group together are specific for this group and Haplotype 3. Based on the Classification Index, the SNPs were not uniformly distributed.

DISCUSSION

The evolution of straining-typing techniques for bacteria have progressed from identifying variations in gel electrophoresis patterns and melt curve analyses to sequencing single pathogen-specific genes and MLST. While typing based on WGS would be ideal, this remains out of reach given the current expense and lack, in general, of sufficient DNA from clinical samples. However, it should be mentioned that we and others have been developing techniques to enrich DNA recovery directly from urogenital and ocular patient sample types with some success (Seth-Smith et al., 2013; Joseph et al., 2014; Hadfield et al., 2017).

*omp*A genotyping remains widely used among *Chlamydia* investigators for molecular epidemiologic and comparative studies of strains between STD and trachoma populations. However, *omp*A encodes for the MOMP, which is under immune selection. MLST offers a more robust typing scheme by employing 6–8 housekeeping genes as relatively immutable signatures for strain typing (Maiden et al., 1998; Dean et al., 2009) and has become an important tool for studying both the epidemiology and evolution of human pathogens (Urwin and Maiden, 2003), including *Ct*.

In this study, we included 58 samples from Russia and eight from Boston, regions that have not previously been represented in *Ct* typing schemes. The 323 reference and clinical samples resolved into 84 STs, representing a 3.23 higher typing resolution over *omp*A genotyping consistent with previous studies (Dean et al., 2009; Gravningen et al., 2012; Batteiger et al., 2014; Herrmann et al., 2015). The high discriminatory index *D* of 0.98 (95% CI 0.97–0.99) and narrow CI for our MLST scheme confirms the validity of this typing method.

We noted an overall high rate of novel STs (67.9%), which may be expected because entire populations were not sampled and the numbers are small for some areas such as Asia and Africa. For example, there were 109 samples from Europe, representing six different countries, and 19 (68%) of the 28 STs were novel. Our findings are similar to other studies in Europe where the rates for novel STs were 62% among high school students in Norway (Gravningen et al., 2012), 65% among Tunisian sex workers (Gharsallah et al., 2016) and 62% among young adults in Amsterdam (Versteeg et al., 2015). Additional novel STs are likely to be identified as new regions undergo MLST. Indeed in

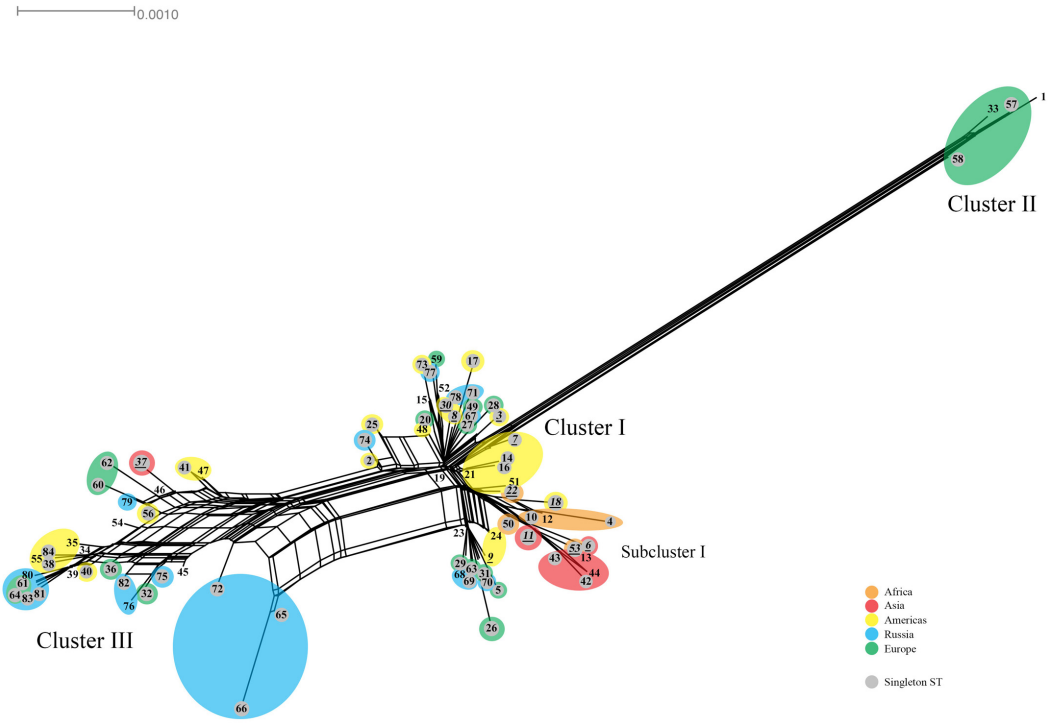


FIGURE 3 | Splitstree for *C. trachomatis* nucleotide sequences by ST. The numbers represent the ST. The tree was obtained by using the seven loci for the 84 ST concatenated nucleotide sequences. STs (shaded gray) are singletons. Orange, Africa; red, Asia; yellow, Americas; blue, Russia; green, Western Europe. Scale bar indicates number of substitutions per site.

Russia, of the 26 STs that contained Russian samples, 18 (69%) were novel.

A significantly higher ST to sample ratio of 0.45 was identified for Russian compared to European and American samples ($p < 0.009$). This was not explained by the number of STs unique to a region as the percentages were relatively uniform across the geographies (**Table 2**). However, Russian women had the highest nucleotide diversity (Pi = 0.00216), followed by North American women (Pi = 0.00210), and Dutch women (Pi = 0.00196), excluding LGV strains that are present only in the female Dutch population and would therefore skew the data. There are no males in the Russian dataset and therefore only women were evaluated here for the three locals. However, when assessing population differentiation for women from the Netherlands, Russia, and North America, excluding those with LGV, there were no significant differences (Supplementary Table 8). A larger sample size for each group would likely provide better resolution of the data. The increased sample ratio for Russia may reflect sexual mixing among the Russian STD cohort with the introduction of STs from other regions or reassortment of existing alleles. Of the 30 alleles present in Russia, 23 (76.7%) were found in other geographic regions, supporting the geographic influx of STs. Reassortment of alleles that generate new STs is also possible given the higher frequency of novel alleles in Russia at 43.3%. It has been shown that recombinational replacements are the major contributors to clonal diversity in contrast to point mutations among bacteria (Spratt et al., 2001). In support of our hypothesis, Hadfield et al. (2017) has shown that *Ct* evolves within genomic 'ecotypes' but also outside of these niches via recombination, consistent with prior genomic studies (Harris et al., 2012; Joseph et al., 2012). However, without a larger sample size from St. Petersburg, the relative overall diversity of STs will remain unknown, and we can only speculate as to the degree of reassortment based on the alleles comprising the STs that represent only the currently sampled cohort of women in Russia.

Excluding singleton STs, STs 23 and 34 were significantly associated with female STD patients in Amsterdam. We had previously noted that *omp*A genotyping is valuable as a separate adjunctive typing method along with MLST as it allows comparison with strains typed only by *omp*A and also can provide evidence for transmission and treatment efficacy as well as putative recombination (Batteiger et al., 2014). But *omp*A should not be included as one of the genes in the MLST scheme because it is under immune selection (Maiden et al., 1998; Dean et al., 2009). We performed *omp*A genotyping on all samples in the database and, for ST23, the samples comprised Ia, B, D, G, H, I, J, and K genotypes while for ST34 they included genotypes F, D, D2, E, J, and Ja (Supplementary Table 3). The high number of *omp*A genotypes for these two STs suggests a high rate of recombination. For example, the sequences of the seven MLST genes matched those of reference strain Ia/UW202 for 47% of the ST23 samples where the *omp*A genotype was also Ia. However, the remaining samples that also matched the seven MLST Ia/UW202

TABLE 6 | *Chlamydia trachomatis* recombinant samples in the dataset.

ST	Number of samples for ST	*ompA* genotype that matches ST sequences	Additional *ompA* genotypes for ST	Number of recombinants (%)*	Geographic region
2	1	D	None	1 (100)	Americas
3	1	H	None	1 (100)	Americas
8	1	K	None	1 (100)	Americas
12	12	A	B, Ba	6 (50)	Tanzania
15	12	K	J	5 (42)	Americas, Russia
16	1	J	None	1 (100)	Americas
17	1	E	None	1 (100)	Americas
19	40	G	B, D, E, H, I, J, K	25 (62.5)	Americas, Europe, Russia
20	1	D	None	1 (100)	Europe
23	32	Ia	A, B, D, G, H, I, J, K	17 (53)	Americas, Europe, Russia
26	1	G	None	1 (100)	Europe
34	38	F	D, E, J, Ja	17 (45)	Americas, Europe
37	1	Da	None	1 (100)	Asia
39	45	E	D, G	2 (4.4)	Americas, Russia
41	1	Ja	None	1 (100)	Americas
45	8	F	D, E	5 (62.5)	Americas, Europe, Russia
46	9	E	D	1 (11)	Europe
49	1	K	None	1 (100)	Europe
50	1	B	None	1 (100)	Africa
53	1	A	None	1 (100)	Africa
60	1	D	None	1 (100)	Europe
62	1	D	None	1 (100)	Europe
65	1	K	None	1 (100)	Russia
66	1	B	None	1 (100)	Russia
68	2	B and G	B and G	2 (100)	Russia
69	1	J	None	1 (100)	Russia
71	1	E	None	1 (100)	Russia
72	1	D	None	1 (100)	Russia
73	1	J	None	1 (100)	Americas
74	1	G	None	1 (100)	Russia
78	1	D	None	1 (100)	Russia
79	2	D and Da	D and Da	2 (100)	Russia
84	1	K	None	1 (100)	Americas

**106 (32.8%) samples were recombinants among the entire database of 323.*

sequences had B, D, G, H, I, J, and K *ompA* genotypes, indicating a mismatch between the MLST and *ompA* sequences, providing evidence for recombination (Supplementary Table 3). Similarly, STs such as 15, 19, and 34 had numerous samples, some of which were putative recombinants. These findings are supported by partial and WGSs where *ompA* has been shown to be involved in frequent exchange and is considered a hotspot for recombination (Gomes et al., 2007; Harris et al., 2012; Joseph et al., 2012). For example, phylogenetic analyses indicate clustering of *ompA* Ja genotypes, which are uncommon, with highly prevalent *ompA*, D, E, and F genotypes (similar to ST34) where hotspots of recombination were noted in *ompA* and *pmpEFGH* genomic regions (Gomes et al., 2007; Joseph et al., 2012). Clusters of *ompA* D genotypes with less prevalent strains G, Ia, and J, similar to our strains in ST23, have also been noted (Harris et al., 2012). In a recent study, *ompA* Ba and C trachoma strains isolated from Australian Aborigines were found to cluster with urogenital D, Da, E, and F strains with hotspots also involving *ompA* and *pmpEFGH* (Andersson et al., 2016). Of course, with WGS, many additional genes have been found to be involved in recombination (Harris et al., 2012; Joseph et al., 2012; Hadfield et al., 2017).

To confirm putative recombinants, we had described that the sequences of the seven MLST genes were individually aligned to the respective gene for all samples in the database and compared these results to the *ompA* genotype of the same sample. For example, **Table 7** shows the five putative recombinants among the eight Boston samples. Sample D/256b had an *ompA* genotype of D but the seven MLST sequences were an exact match to the seven sequences of F. In some cases, one or more of the seven genes matched two different strains as was the case for samples J/253b and J/259b. Other examples of recombinants are shown in Supplementary Tables 6 and 7 for the Dutch and Russian samples, respectively. In addition, Splitstree analysis was performed and revealed ancillary evidence for a network structure consistent with homologous recombination

TABLE 7 | Evidence for recombination among the Boston, MA, United States *C. trachomatis* samples.

Sample	*omp*A genotype	ST sequence homology*	ST	Location of allelic SNPs	
J/253b	**J**	**G or K**	**15 (G/K)**	**G: *pdhA*: 339 C→T *lysS*: 34 A→G**	**K: *pykF*: 40 G→A *lysS*: 34 A→G**
Ia/258b	Ia	Ia	23 (Ia)		
F/255b	F	F	34 (F)		
D/256b	**D**	**F**	**34 (F)**		
D/257b	**D**	**E**	**39 (E)**		
E/260b	E	E	39 (E)		
J/259b	**J**	**G or K**	**73 (G/K)**	**G: *pdhA*: 339 C→T *yhbG*: 433 G→C *lysS*: 34 A→G *leuS*: 282 T→C**	**K: *yhbG*: 433 G→C *pykF*: 40 G→A *lysS*: 34 A→G *leuS*: 282 T→C**
K/254b	**K**	**F**	**84 (F)**		

*Bold denotes putative recombinant. *The seven ST genes of the sample have the highest homology to the seven genes of the strain that has the ompA genotype denoted in the column (e.g., for sample D/256b that has a D ompA genotype, the seven ST genes were identical to the seven genes of F strains in the database where the ompA genotype was also F for those strains).*

TABLE 8 | Single nucleotide polymorphisms (SNP) correlate with haplotype and disease phenotype.

Gene Locus	Number of SNPs per locus	Haplotype 1 All clinical and reference invasive LGV strains	Haplotype 2 Prevalent reference and clinical Da, E, F, and Ja strains (except clinical D/83s, D84s, D2s, D/EC, D/LC, D/SotonD5, D/SotonD6, D43nl, D/202nl, D/203nl, D/204nl, D/205nl, D/206nl, D/207nl, E87e, and reference D/UW3/Cx)	Haplotype 3 All reference and clinical trachoma A, B, Ba, and C strains (except reference strain A/Sa-1 and strain B/Jali20)
		Disease phenotype: LGV*	Disease phenotype: urogenital non-invasive*	Disease phenotype: trachoma*
glyA	1–7			
fmdhC	1–5			
pdhA	1–6			
yhbG	1–21	SNPs 2–6	SNPs 15	
		8–12	18	
		16	20	
pykF	1–9		SNPs 6	SNP 3
			7	
lysS	1–11			
leuS	1–12	SNPs 1		SNP 3
		5		
		11		
Total	**71**	**14**	**5**	**2**

*LGV, lymphogranuloma venereum. *$p < 0.01$.*

(**Figure 2**). While these data confirm the recombinant nature of the samples, it is likely that there are other genetic regions that have undergone recombination and, therefore, it is not possible to determine the extent of genetic exchange unless WGS is performed (Joseph et al., 2011, 2012; Harris et al., 2012; Hadfield et al., 2017).

Overall, there were 106 (32.8%) putative recombinants (**Table 2** and Supplementary Table 3), which is similar to our previous studies and those of other investigators (Dean et al., 2009; Gravningen et al., 2012; Batteiger et al., 2014). Each geographic region contained recombinants, although Europe had a significantly higher number than Russia and the Americas ($p < 0.04$) (**Table 2**). This result was skewed by the higher rate of recombinants among the Amsterdam population, which would be expected given that the samples came from individuals at high risk for STDs where sexual mixing and import of strains from tourists could increase the chances for multiple *Ct* strain infections and opportunities for recombination. Indeed, rates of *Ct* mixed infections as high as 6 to 16% have been reported among men who have sex with men and heterosexual populations, respectively, in Europe, including the Netherlands (Quint et al., 2011; Rodriguez-Dominguez et al., 2015).

The most geographically prevalent ST was 39 with 45 samples, 95% of which had an *omp*A genotype of E; there were only two recombinants in this ST: one from Boston with a D *omp*A genotype and one from St. Petersburg with a G genotype. E genotypes are known to be the most globally prevalent (Lysén et al., 2004; Millman et al., 2004; Spaargaren et al., 2004;

Lee et al., 2006; Gharsallah et al., 2016) and the least recombinogenic based on whole genome sequencing (Joseph et al., 2011, 2012). In our dataset, there were 72 E genotypes with an ST to sample ratio of 0.25; only the F genotype, the 4th most prevalent genotype with 30 samples, had a lower ratio at 0.17. The lower ratios indicate greater fitness as these strains are prevalent worldwide and have fewer allelic variants that resolve into fewer STs. This is borne out by the fact that only eight (11%) of the 72 E genotypes and 0 of the 30 F genotypes were recombinants (Supplementary Table 3). A recent genome study that included 149 E strains supports our conclusions (Hadfield et al., 2017). Genotype D was also highly prevalent but had a much higher ratio at 0.48, and 29 (88%) of 33 samples were recombinants.

We previously found that phylogeny based on MLST resolved the STs along disease phenotype demarcations. As samples have been added to the database, the phenotype resolution has increased to include the LGV phenotype, denoted as clonal cluster-D (CC-D; **Figure 1**). Similarly, the tree shows the three main clusters with the trachoma STs as a Subcluster of Cluster I, which resembles those of other reports (Herrmann et al., 2015).

To determine whether the phenotypic groups could be more finely discriminated, we analyzed the database for SNPs that independently or together would identify a phenotype. As in our previous studies (Dean et al., 2009; Batteiger et al., 2014), specific SNPs correlated with LGV, non-invasive urogenital disease and trachoma. Haplotype 2, which included the non-invasive urogenital disease group required exclusion of strains that were recombinants, specifically D genotypes.

AUTHOR CONTRIBUTIONS

Conceived and designed experiments: DD and VS. Wrote the paper: VS, AV, and DD. Assisted with editing the paper: EFvE, MPN, LAS, HJCdV, and SAM. Performed experiments: EFvE, MPN, TM, and BS. Analyzed data: VS, AV, EFvE, MPN, TM, BS, RW, and DD. Contributed reagents/materials/analysis tools: VS, LAS, HJCdV, JD, CE, SAM, and DD.

FUNDING

This work was funded in part by National Institutes of Health grants R01 AI 098843 to DD and R03 TW 007754 to DD.

ACKNOWLEDGMENT

We would like to thank Payel Sen, Neda Dastgheyb, and Priyanka Manandhar for excellent technical support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02195/full#supplementary-material

REFERENCES

Andersson, P., Harris, S. R., Seth Smith, H. M., Hadfield, J., O'Neill, C., Cutcliffe, L. T., et al. (2016). *Chlamydia trachomatis* from Australian Aboriginal people with trachoma are polyphyletic composed of multiple distinctive lineages. *Nat. Comm.* 7:10688. doi: 10.1038/ncomms10688

Batteiger, B. E., Wan, R., Williams, J. A., He, L., Ma, A., Fortenberry, J. D., et al. (2014). Novel *Chlamydia trachomatis* strains in heterosexual sex partners, Indianapolis, Indiana, USA. *Emerg. Infect. Dis.* 20, 1841–1847. doi: 10.3201/2011.140604

Baud, D., Regan, L., and Greub, G. (2008). Emerging role of Chlamydia and Chlamydia-like organisms in adverse pregnancy outcomes. *Curr. Opin. Infect. Dis.* 21, 70–76. doi: 10.1097/QCO.0b013e3282f3e6a5

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–300. doi: 10.2307/2346101

Blas, M. M., Canchihuaman, F. A., Alva, I. E., and Hawes, S. E. (2007). Pregnancy outcomes in women infected with *Chlamydia trachomatis*: a population-based cohort study in Washington State. *Sex. Transm. Infect.* 83, 314–318. doi: 10.1136/sti.2006.022665

Borkowf, C. B. (2006). Constructing binomial confidence intervals with near nominal coverage by adding a single imaginary failure or success. *Stat. Med.* 25, 3679–3695. doi: 10.1002/sim.2469

Danby, C. S., Cosentino, L. A., Rabe, L. K., Priest, C. L., Damare, K. C., Macio, I. S., et al. (2016). Patterns of extragenital Chlamydia and Gonorrhea in women and men who have sex with men reporting a history of receptive anal intercourse. *Sex. Transm. Dis.* 43, 105–109. doi: 10.1097/OLQ.0000000000000384

Darville, T. (2005). *Chlamydia trachomatis* infections in neonates and young children. *Semin. Pediatr. Infect. Dis.* 16, 235–244. doi: 10.1053/j.spid.2005.06.004

Dean, D. (2010). "Pathogenesis of chlamydial ocular infections," in *Duane's Foundations of Clinical Ophthalmology*, eds W. Tasman and E. A. Jaeger (Philadelphia, PA: Lippincott Williams & Wilkins), 678–702. "Chapter 77.

Dean, D., Bruno, W. J., Wan, R., Gomes, J. P., Devignot, S., Mehari, T., et al. (2009). Predicting phenotype and emerging strains among *Chlamydia trachomatis* infections. *Emerg. Infect. Dis.* 15, 1385–1394. doi: 10.3201/eid1509.090272

Dean, D., Kandel, R. P., Adhikari, H. K., and Hessel, T. (2008). Multiple Chlamydiaceae species in trachoma: implications for disease pathogenesis and control. *PLOS Med.* 5:e14. doi: 10.1371/journal.pmed.0050014

Dean, D., and Millman, K. (1997). Molecular and mutation trends analyses of omp1 alleles for serovar E of *Chlamydia trachomatis*. Implications for the immunopathogenesis of disease. *J. Clin. Invest.* 99, 475–483. doi: 10.1172/JCI119182

Dean, D., Schachter, J., Dawson, C. R., and Stephens, R. S. (1992). Comparison of the major outer membrane protein variant sequence regions of B/Ba isolates: a molecular epidemiologic approach to *Chlamydia trachomatis* infections. *J. Infect. Dis.* 166, 383–392. doi: 10.1093/infdis/166.2.383

de Waaij, D. J., Dubbink, J. H., Peters, R. P., Ouburg, S., and Morré, S. A. (2015). Comparison of GMT presto assay and Roche cobas 4800 CT/NG assay for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in dry swabs. *J. Microbiol. Methods* 118, 70–74. doi: 10.1016/j.mimet.2015.08.020

Excoffier, L., and Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10, 564–567. doi: 10.1111/j.1755-0998.2010.02847.x

Feil, E. J., Li, B. C., Aanensen, D. M., Hanage, W. P., and Spratt, B. G. (2004). eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J. Bacteriol.* 186, 1518–1530. doi: 10.1128/JB.186.5.1518-1530.2004

Gharsallah, H., Bom, R. J. M., Bruisten, S. M., Himschoot, M., Frikha-Gargouri, O., and Hammami, A. (2016). Identification of a dominant *Chlamydia trachomatis* strain in patients attending sexual transmitted infection clinic and female sex

workers in Tunisia using a high resolution typing method. *Infect. Genet. Evol.* 44, 444–449. doi: 10.1016/j.meegid.2016.08.002

Gomes, J. P., Bruno, W. J., Borrego, M. J., and Dean, D. (2004). Recombination in the genome of *Chlamydia trachomatis* involving the polymorphic membrane protein C relative to *ompA* and evidence for horizontal gene transfer. *J. Bacteriol.* 186, 4295–4306. doi: 10.1128/JB.186.13.4295-4306.2004

Gomes, J. P., Bruno, W. J., Borrego, M. J., Nunes, A., Florindo, C., and Dean, D. (2006). Polymorphisms in the nine polymorphic membrane proteins of *Chlamydia trachomatis* across all serovars: evidence for Da recombination and correlation with tissue tropism. *J. Bacteriol.* 188, 275–286. doi: 10.1128/JB.188.1.275-286.2006

Gomes, J. P., Bruno, W. J., Nunes, A., Santos, N., Florindo, C., Borrego, M. J., et al. (2007). Evolution of *Chlamydia trachomatis* diversity occurs by widespread interstrain recombination involving hotspots. *Genome Res.* 17, 50–60. doi: 10.1101/gr.5674706

Götz, H. M., Bom, R. J. M., Wolfers, M. E. G., Fennema, J., van den Broek, I. V. F., Speksnijder, A. G. C. L., et al. (2013). Use of *Chlamydia trachomatis* high-resolution typing: an extended case study to distinguish recurrent or persistent infection from new infection. *Sex. Transm. Infect.* 90, 155–160. doi: 10.1136/sextrans-2013-051218

Gravningen, K., Christerson, L., Furberg, A. S., Simonsen, G. S., Ödman, K., Ståhlsten, A., et al. (2012). Multilocus sequence typing of genital *Chlamydia trachomatis* in Norway reveals multiple new sequence types and a large genetic diversity. *PLOS ONE* 7:e34452. doi: 10.1371/journal.pone.0034452

Hadfield, J., Harris, S. R., Seth-Smith, H. M. B., Parmar, S., Andersson, P., Giffard, P. M., et al. (2017). Comprehensive global genome dynamics of *Chlamydia trachomatis* show ancient diversification followed by contemporary mixing. *Genome Res.* 7, 1220–1229. doi: 10.1101/gr.212647.116

Harris, S. R., Clarke, I. N., Seth-Smith, H. M., Solomon, A. W., Cutcliffe, L. T., Marsh, P., et al. (2012). Whole-genome analysis of diverse *Chlamydia trachomatis* strains identifies phylogenetic relationships masked by current clinical typing. *Nat. Genet.* 44, 413–419, S1. doi: 10.1038/ng.2214

Herrmann, B., Isaksson, J., Ryberg, M., Tångrot, J., Saleh, I., Versteeg, B., et al. (2015). Global multilocus sequence type analysis of *Chlamydia trachomatis* strains from 16 countries. *J. Clin. Microbiol.* 53, 2172–2179. doi: 10.1128/JCM.00249-15

Horn, M., Collingro, A., Schmitz-Esser, S., Beier, C. L., Purkhold, U., Fartmann, B., et al. (2004). Illuminating the evolutionary history of chlamydiae. *Science* 304, 728–730. doi: 10.1126/science.1096330

Hunter, P. R., and Gaston, M. A. (1988). Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J. Clin. Microbiol.* 26, 2465–2466.

Isaksson, J., Gallo Vaulet, L., Christerson, L., Ruettger, A., Sachse, K., Entrocassi, C., et al. (2016). Comparison of multilocus sequence typing and multilocus typing microarray of *Chlamydia trachomatis* strains from Argentina and Chile. *J. Microbiol. Methods* 127, 214–218. doi: 10.1016/j.mimet.2016.06.005

Joseph, S. J., Didelot, X., Gandhi, K., Dean, D., and Read, T. D. (2011). Interplay of recombination and selection in the genomes of *Chlamydia trachomatis*. *Biol. Direct* 6:28. doi: 10.1186/1745-6150-6-28

Joseph, S. J., Didelot, X., Rothschild, J., de Vries, H. J. C., Morré, S. A., Read, T. D., et al. (2012). Population genomics of *Chlamydia trachomatis*: insights on drift, selection, recombination, and population structure. *Mol. Biol. Evol.* 29, 3933–3946. doi: 10.1093/molbev/mss198

Joseph, S. J., Li, B., Ghonasgi, T., Qin, Z. C., Dean, D., and Read, T. D. (2014). Direct amplification, sequencing and profiling of *Chlamydia trachomatis* strains in single and mixed infection clinical samples. *PLOS ONE* 9:e99290. doi: 10.1371/journal.pone.0109670

Klint, M., Fuxelius, H. H., Goldkuhl, R. R., Skarin, H., Rutemark, C., Andersson, S. G. E., et al. (2007). High-resolution genotyping of *Chlamydia trachomatis* strains by multilocus sequence analysis. *J. Clin. Microbiol.* 45, 1410–1414. doi: 10.1128/JCM.02301-06

Labiran, C., Marsh, P., Zhou, J., Bannister, A., Clarke, I. N., Goubet, S., et al. (2016). Highly diverse MLVA-*omp*A genotypes of rectal *Chlamydia trachomatis* among men who have sex with men in Brighton, UK and evidence for an HIV-related sexual network. *Sex. Transm. Infect.* 92, 299–304. doi: 10.1136/sextrans-2015-052261

Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187

Lee, G., Park, J., Kim, B., Kim, S.-A., Yoo, C.-K., and Seong, W. K. (2006). *OmpA* genotyping of *Chlamydia trachomatis* from Korean female sex workers. *J. Infect.* 52, 451–454. doi: 10.1016/j.jinf.2005.08.019

Lysén, M., Osterlund, A., Rubin, C.-J., Persson, T., Persson, I., and Herrmann, B. (2004). Characterization of *ompA* genotypes by sequence analysis of DNA from all detected cases of *Chlamydia trachomatis* infections during 1 year of contact tracing in a Swedish County. *J. Clin. Microbiol.* 42, 1641–1647. doi: 10.1128/JCM.42.4.1641-1647.2004

Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., et al. (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. U.S.A.* 95, 3140–3145. doi: 10.1073/pnas.95.6.3140

Mårdh, P.-A. (2004). Tubal factor infertility, with special regard to chlamydial salpingitis. *Curr. Opin. Infect. Dis.* 17, 49–52. doi: 10.1097/00001432-200402000-00010

Millman, K., Black, C. M., Johnson, R. E., Stamm, W. E., Jones, R. B., Hook, E. W., et al. (2004). Population-based genetic and evolutionary analysis of *Chlamydia trachomatis* urogenital strain variation in the United States. *J. Bacteriol.* 186, 2457–2465. doi: 10.1128/JB.186.8.2457-2465.2004

Millman, K., Tavere, S., and Dean, D. (2001). Recombination in the *ompA* gene but not the omcB gene of Chlamydia contribute to serovar specific differences in tissue tropism, immune surveillance and persistence of the organism. *J. Bacteriol.* 183, 5997–6008. doi: 10.1128/JB.183.20.5997-6008.2001

Morré, S. A., Van Valkengoed, I. G., Moes, R. M., Boeke, A. J., Meijer, C. J., and Van den Brule, A. J. (1999). Determination of *Chlamydia trachomatis* prevalence in an asymptomatic screening population: performances of the LCx and COBAS Amplicor tests with urine specimens. *J. Clin. Microbiol.* 37, 3092–3096.

Pannekoek, Y., Morelli, G., Kusecek, B., Morré, S. A., Ossewaarde, J. M., Langerak, A. A., et al. (2008). Multi locus sequence typing of *Chlamydiales*: clonal groupings within the obligate intracellular bacteria *Chlamydia trachomatis*. *BMC Microbiol.* 8:42. doi: 10.1186/1471-2180-8-42

Perkins, E., and Hill, D. (eds) (2013). *Scientific Foundations of Ophthalmology*. Oxford: Butterworth-Heinemann.

Peuchant, O., Touati, A., Sperandio, C., Hénin, N., Laurier-Nadalié, C., Bébéar, C., et al. (2016). Changing pattern of *Chlamydia trachomatis* strains in Lymphogranuloma Venereum outbreak, France, 2010-2015. *Emerg. Infect. Dis.* 22, 1945–1947. doi: 10.3201/eid2211.160247

Quint, K. D., Bom, R. J., Quint, W. G. V., Bruisten, S. M., van der Loeff, M. F. S., Morré, S. A., et al. (2011). Anal infections with concomitant *Chlamydia trachomatis* genotypes among men who have sex with men in Amsterdam, the Netherlands. *BMC Infect. Dis.* 11:63. doi: 10.1186/1471-2334-11-63

Rodriguez-Dominguez, M., Gonzalez-Alba, J. M., Puerta, T., Menendez, B., Sanchez-Diaz, A. M., Canton, R., et al. (2015). High prevalence of co-infections by invasive and non-invasive *Chlamydia trachomatis* genotypes during the Lymphogranuloma Venereum outbreak in Spain. *PLOS ONE* 10:e0126145. doi: 10.1371/journal.pone.0126145

Rowley, J., Toskin, I., Ndowa, F., World Health Organization, and Reproductive Health and Research. (2012). *Global Incidence and Prevalence of Selected Curable Sexually Transmitted Infections, 2008*. Geneva: World Health Organization.

Schliep, K. P. (2010). phangorn: phylogenetic analysis in R. *Bioinformatics* 27, 592–593. doi: 10.1093/bioinformatics/btq706

Schillinger, J. A., Katz, B. P., Markowitz, L. E., Braslins, P. G., Shrier, L. A., Madico, G., et al. (2016). Genotype-specific concordance of *Chlamydia trachomatis* genital infection within heterosexual partnerships. *Sex. Transm. Dis.* 43, 741–749. doi: 10.1097/OLQ.0000000000000525

Sethi, G., Allason-Jones, E., Richens, J., Annan, N. T., Hawkins, D., Ekbote, A., et al. (2009). Lymphogranuloma venereum presenting as genital ulceration and inguinal syndrome in men who have sex with men in London, UK. *Sex. Transm. Infect.* 85, 165–170. doi: 10.1136/sti.2008.034348

Seth-Smith, H. M., Harris, S. R., Skilton, R. J., Radebe, F. M., Golparian, D., Shipitsyna, E., et al. (2013). Whole-genome sequences of *Chlamydia*

trachomatis directly from clinical samples without culture. *Genome Res.* 5, 855–866. doi: 10.1101/gr.150037.112

Shalepo, K., Savicheva, A., Shipitsyna, E., Unemo, M., and Domeika, M. (2006). Diagnosis of *Chlamydia trachomatis* in Russia–in-house PCR assays may be effective but overall optimization and quality assurance are urgently needed. *APMIS Acta Pathol. Microbiol. Immunol. Scand.* 114, 500–507. doi: 10.1111/j.1600-0463.2006.apm_443.x

Shipitsyna, E., Shalepo, K., Savicheva, A., Unemo, M., and Domeika, M. (2007). Pooling samples: the key to sensitive, specific and cost-effective genetic diagnosis of *Chlamydia trachomatis* in low-resource countries. *Acta Derm. Venereol.* 87, 140–143. doi: 10.2340/00015555-0196

Simpson, E. H. (1949). Measurement of Diversity. *Nature* 163:688. doi: 10.1038/163688a0

Smelov, V., Quint, K. D., Pleijster, J., Savelkoul, P. H. M., Shalepo, K., Shipitsyna, E., et al. (2009). *Chlamydia trachomatis* serovar distributions in Russian men and women: a comparison with Dutch serovar distributions. *Drugs Today* 45, Suppl B:33–38.

Spaargaren, J., Verhaest, I., Mooij, S., Smit, C., Fennema, H. S. A., Coutinho, R. A., et al. (2004). Analysis of *Chlamydia trachomatis* serovar distribution changes in the Netherlands (1986-2002). *Sex. Transm. Infect.* 80, 151–152. doi: 10.1136/sti.2003.006395

Spratt, B. G., Hanage, W. P., and Feil, E. J. (2001). The relative contributions of recombination and point mutation to the diversification of bacterial clones. *Curr. Opin. Microbiol.* 4, 602–606. doi: 10.1016/S1369-5274(00)00257-5

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Urwin, R., and Maiden, M. C. J. (2003). Multi-locus sequence typing: a tool for global epidemiology. *Trends Microbiol.* 11, 479–487. doi: 10.1016/j.tim.2003.08.006

van Belkum, A., Tassios, P. T., Dijkshoorn, L., Haeggman, S., Cookson, B., Fry, N. K., et al. (2007). Guidelines for the validation and application of typing methods for use in bacterial epidemiology. *Clin. Microbiol. Infect.* 13(Suppl. 3), 1–46. doi: 10.1111/j.1469-0691.2007.01786.x

Versteeg, B., Himschoot, M., van den Broek, I. V. F., Bom, R. J. M., Speksnijder, A. G. C. L., van der Loeff, M. F., et al. (2015). Urogenital *Chlamydia trachomatis* strain types, defined by high-resolution multilocus sequence typing, in relation to ethnicity and urogenital symptoms among a young screening population in Amsterdam, The Netherlands. *Sex. Transm. Infect.* 91, 415–422. doi: 10.1136/sextrans-2014051790

WHO Sexually transmitted infections [STIs] (2015). Available at: http://www.who.int/mediacentre/factsheets/fs110/en/ [Accessed June 5, 2015].

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplemental Material

Chlamydia trachomatis strain types have diversified regionally and globally with evidence for recombination across geographic divides

Vitaly Smelov[1,2,3]†, Alison Vrbanac[4]†§, Eleanne F. van Ess[5]†, Marlies P. Noz[5], Raymond Wan[4], Carina Eklund[2], Tyler Morgan[4], Lydia A. Shrier[6], Blake Sanders[4], Joakim Dillner[2], Henry J. C. de Vries[7,8.9], Servaas A. Morre[5,10], and Deborah Dean[4,11,12*]

[1]International Agency for Research on Cancer, World Health Organization, Lyon, France, [2]Karolinska Institute, Stockholm, Sweden, [3]North-Western State Medical University named after I.I. Mechnikov, St. Petersburg, Russia, [4]UCSF Benioff Children's Hospital Oakland Research Institute, Oakland, CA, United States, [5]Laboratory of Immunogenetics, Department of Medical Microbiology and Infection Control, VU University Medical Center, Amsterdam, the Netherlands, [6]Department of Pediatrics, Boston Children's Hospital, Boston, Massachusetts, United States, [7]Center for Infection and Immunology Amsterdam (CINIMA), Academic Medical Center, 1105 AZ, Amsterdam, The Netherlands, [8]Department of Dermatology, Academic Medical Center, 1105 AZ, Amsterdam, The Netherlands, [9]STI Outpatient Clinic, Public Health Service of Amsterdam (GGD Amsterdam), 1018 WT, Amsterdam, The Netherlands, [10]Institute of Public Health Genomics, Department of Genetics and Cell Biology, Research Institute GROW (School for Oncology & Developmental Biology), Faculty of Health, Medicine & Life Sciences, University of Maastricht, Maastricht, The Netherlands, [11]Department of Bioengineering, University of California at Berkeley and University of California at San Francisco, CA, United States, [12]Departments of Medicine and Pediatrics, UCSF, San Francisco, CA, United States

Tables

Supplementary Table 1. Primers used for PCR and Sequencing.

Supplementary Table 2. Characteristics of Alleles for Each Locus.

Supplementary Table 3. Sequence types, allelic profiles, and clinical characteristics of reference and clinical samples.

Supplementary Table 4: eBURST report for 323 isolates.

Supplementary Table 5: STs by geographic region.

Supplemental Table 6: Evidence of recombination for samples from Amsterdam, The Netherlands.

Supplemental Table 7: Evidence of recombination for samples from St. Petersburg, Russia.

Supplementary Table 8: Pairwise Population Differentiation (Fst) for North American, Dutch, and Russian Women**.**

Figures

Supplementary Figure 1: Minimum evolution tree for *C. trachomatis* amino acid sequences by ST. The numbers represent the Sequence Type (ST). The tree was constructed using the matrix of pairwise differences between the 84 ST concatenated amino acid sequences for the seven loci using the maximum composite likelihood method for estimating genetic distances. Only bootstrap values (1,000 replicates) >70% are represented at branch points in the tree. STs shaded grey are singleton STs. The area of the circle represents the number of samples of each ST. Orange, Africa; red, Asia; yellow, The America; blue, Russia; green, Western Europe. Scale bar indicates number of substitutions per site.

Supplementary Figure 2: Splitstree for *C. trachomatis* amino acid sequences by ST. The numbers represent the Sequence Type (ST). The tree was obtained by using the 7 loci for the 84 ST concatenated sequences. STs shaded grey are singleton STs. The area of the circle represents the number of samples of each ST. Orange, Africa; red, Asia; yellow, The America; blue, Russia; green, Western Europe. Scale bar indicates number of substitutions per site.

Supplementary Table 1: Primers used for PCR and Sequencing

Target Locus	Region	PCR Primer Label	Sequence 5' to 3'	Sequencing Primer Label	Sequence 5' to 3' [2]	Length of Sequence, bp
glyA	CT432	1FglyA	GAAGACTGTGGCGCTGTT	G1FSP	GCTCATTTCGCAGGCTTGGTT	523
		1RglyA	CTTCCTGAGCGATCCCTTC	G1RSP	GACAGCTCTCCCTCACTTT	
		Alternates:		1FglyA	GAAGACTGTGGCGCTGTTT	
		G1PCRFa	GAACATAAGCCCACCGTTCT	1RglyA	CTTCCTGAGCGATCCCTTC	
		G1PCRRa	TTCCAGATCGATTTCAGGAT			
mdhC	CT376	3FmdhC	GGAGATGTTTTTGGCCTTG	G3FSP	GATGTTTTTGGCCTTGATTG	521
		3RmdhC	CGATTACTGCACTACCACG	G3RSP	CTGTACAGAAGGCACCATAATA	
		Alternates:		3FmdhC	GGAGATGTTTTTGGCCTTG	
		G3PCRFa	AGGGCAAATAGCCTATAGCT	3RmdhC	CGATTACTGCACTACCACG	
		G3PCRRa	AAGCTCGTGCTGCAGAAGCT			
pdhA	CT245	5FpdhA	CTACAGAAGCCCGAGTTTT	G5FSP	ATTCTTTCTGCATTGACCT	551
		5RpdhA	CTGTTTGTTGCATGTGGTG	G5RSP	CATGTGGTGATAAGCTTCTCTAAA	
		Alternate:				
		G5PCRFa	CATCCTCTGACTCTCAACAT	5FpdhA	CTACAGAAGCCCGAGTTTT	
		G5PCRRa	TAGGATCGGAAATAGAGTGT	5RpdhA	CTGTTTGTTGCATGTGGTG	
ybhG	CT653	8FyhbG	TCAAGTCAATGCAGGAGAA	G8FSP	CCTAACGGAGCGGGGAAA	504
		8RybhG	GATAGTGTTGACGTACCATAGGAT	G8RSP	GAGCAGACGATCCTTCAA	
pykF	CT332	9FpykF	ATCTTATCGCTGCTTCGTT	G9SeqF	TGTAATGAAGATATTGACA	527
		9RpykF	CAGCAATAATAGGGAGATA	G9SeqR	TTGGATAAAAACATCGGAG	
		Alternates:				
		G9seqF	TGTAATGAAGATATTGACAGCAT	G9SeqF2	GATATTGACAGCATGCGTAAAGTT	
		G9seqR	TTGGATAAAAACATCGGAG	G9SeqR2	ATTGTGTATACCCAGACGGGAGGAT	
lysS	CT781	11FPCR2	GAATGTCCCGAGTTTATGAA	11FlysS	GAAGGAATCGATAGAACGCATAAT	577
		11RPCR2	GTATAGAAGCAAAAAAAGAT	11RlysS	ATACGCCGCATAACAGGGAAAAAC	
		Alternates:				
		G11FSP	TCCAGAGTTTACCATGATAGAGG	11seqF	TATGCGGCTTATTGGGATTA	
		G11RSP	AACAGGGAAAAACAGGACAT	11seqR	AATGGATGCGGCGTCTGTCAA	
leuS	CT209	12F5	GAGCTGGATTGGCCAGAAAGT	12DeanF5	TCTTCAGGAGCTTCTGTTAATT	521
		12R5	GCCAGAAACGCGCATAAAGT	G12PCRRa	AGTGCAGTACAGCATGTTCT	
				G12RSP	AAGAGCCTGCCCATTGAG	
		Alternates:				
		G12PCRFa	ACAAGACCGGACACTTTGAT	G12PCRFa	ACAAGACCGGACACTTTGAT	
		G12PCRRa	AGTGCAGTACAGCATGTTCT	G12RSP	CTCAATGGGCAGGCTCTT	

Supplementary Table 2. Characteristics of alleles for each locus

Gene locus	No. alleles	Length, bp	No. polymorphic sites	Average pairwise distance	Average dS	Average dN
glyA	12	523	9	0.0032	0.0019	0.0013
mdhC	7	521	6	0.0013	0.0009	0.0004
pdhA	8	551	6	0.0002	0.0001	0.0001
yhbG	11	504	22	0.0109	0.0100	0.0009
pykF	10	527	13	0.0029	0.0013	0.0016
lysS	12	577	15	0.0015	0.0011	0.0004
leuS	18	521	19	0.0024	0.0007	0.0017
Total	78	3,724	90	0.0224	0.0160	0.0064

Supplementary Table 3. Sequence types, allelic profiles, and clinical characteristics of reference and clinical samples*

		Allele assignment for each locus								
Strain ID †	ST	*glyA*	*mdhC*	*pdhA*	*yhbG*	*pykF*	*lysS*	*leuS*	Region of isolation	Diagnosis/Site
L1/440	1	1	1	3	8	1	4	11	California	LGV
L2/434	1	"	"	"	"	"	"	"	California	LGV
L3/404	1	"	"	"	"	"	"	"	California	LGV
L2b/Canada1	1	"	"	"	"	"	"	"	Canada	Proctitis
L2b/Canada2	1	"	"	"	"	"	"	"	Canada	Proctitis
L2b/CV204	1	"	"	"	"	"	"	"	France	Proctitis
L2b/LST	1	"	"	"	"	"	"	"	France	Proctitis
L2b/795	1	"	"	"	"	"	"	"	France	Proctitis
L2/54s	1	"	"	"	"	"	"	"	San Francisco	Proctitis
L2a/UW396	1	"	"	"	"	"	"	"	Seattle	LGV
L1/115	1	"	"	"	"	"	"	"	South Africa	LGV
L1/1322/p2	1	"	"	"	"	"	"	"	South Africa	LGV
L1/224	1	"	"	"	"	"	"	"	South Africa	LGV
L2b/8200/07	1	"	"	"	"	"	"	"	Sweden	Proctitis
L2b/Ams1	1	"	"	"	"	"	"	"	The Netherlands	Proctitis
L2b/Ams2	1	"	"	"	"	"	"	"	The Netherlands	Proctitis
L2b/Ams3	1	"	"	"	"	"	"	"	The Netherlands	Proctitis
L2b/Ams4	1	"	"	"	"	"	"	"	The Netherlands	Proctitis
L2b/Ams5	1	"	"	"	"	"	"	"	The Netherlands	Proctitis
L2b/86nl	1	"	"	"	"	"	"	"	The Netherlands	Proctitis
L2/237nl	1	"	"	"	"	"	"	"	The Netherlands	Cervix
L2b/UCH1	1	"	"	"	"	"	"	"	United Kingdom	Proctitis
L2b/UCH2	1	"	"	"	"	"	"	"	United Kingdom	Proctitis
L2/25667R	1	"	"	"	"	"	"	"	USA	Proctitis
L2c	1	"	"	"	"	"	"	"	San Francisco	Proctitis
D/84s	**2**	**2**	**3**	**3**	**6**	**5**	**4**	**3**	**San Francisco**	**Cervicitis**
H/UW4/Cx	**3**	**3**	**1**	**3**	**6**	**6**	**4**	**3**	**Washington**	**Cervicitis**
A/51t	4	3	3	1	6	3	7	9	Tanzania	Trachoma
H/46nl	5	3	3	2	6	6	8	3	The Netherlands	Cervicitis and vaginal discharge
B/TW5/OT	6	3	3	3	4	3	5	10	Taiwan	Conjunctivitis
I/UW12/Ur	7	3	3	3	6	1	4	3	Washington	Urethritis
K/UW36/Cx	**8**	**3**	**3**	**3**	**6**	**2**	**4**	**3**	**Washington**	**Cervicitis**
J/UW36/Cx	9	3	3	3	6	2	8	3	Washington	Cervicitis
Ja/UW92	9	"	"	"	"	"	"	"	Washington	Cervicitis
B/53t	10	3	3	3	6	3	4	9	Tanzania	Trachoma
C/TW3/OT	11	3	3	3	6	3	5	7	Taiwan	Conjunctivitis
A/2497	12	3	3	3	6	3	5	9	Tanzania	Trachoma
A/363	12	"	"	"	"	"	"	"	Tanzania	Trachoma
A/48t	12	"	"	"	"	"	"	"	Tanzania	Trachoma
A/5291	12	"	"	"	"	"	"	"	Tanzania	Trachoma
A/59t	12	"	"	"	"	"	"	"	Tanzania	Trachoma
A/7249	12	"	"	"	"	"	"	"	Tanzania	Trachoma
B/TZ1A828/OT	**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	3	3	3	6	3	6	7	Nepal	Trachoma, TS
C/33n	13	"	"	"	"	"	"	"	Nepal	Trachoma, TS
G/15s	14	3	3	3	6	4	4	8	San Francisco	Proctitis
K/186i	15	3	3	3	6	6	1	6	Indianapolis	Urethra
K/187i	15	"	"	"	"	"	"	"	Indianapolis	Cervix
K/296sp	15	"	"	"	"	"	"	"	St. Petersburg	Cervicitis
K/305sp	15	"	"	"	"	"	"	"	St. Petersburg	Cervicitis

K/42nl	15	“	“	“	“	“	“	“	The Netherlands	Cervicitis w/ vaginal discharge
K/49nl	15	“	“	“	“	“	“	“	The Netherlands	Cervicitis w/ vaginal discharge
K/250nl	15	“	“	“	“	“	“	“	The Netherlands	Cervix
J/253b	**15**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**Boston**	**Cervicitis**
J/112i	**15**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**Indianapolis**	**Urethra**
J/113i	**15**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**Indianapolis**	**Cervix**
J/31-98	**15**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**Seattle**	**Cervicitis**
J/318sp	**15**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**St. Petersburg**	**Cervicitis**
J/27s	**16**	**3**	**3**	**3**	**6**	**6**	**1**	**8**	**San Francisco**	**Cervicitis/urethritis**
E/87e	**17**	**3**	**3**	**3**	**6**	**6**	**2**	**3**	**Ecuador**	**Cervicitis**
Ba/Apache2	18	3	3	3	6	6	3	9	Arizona	Conjunctivitis
G/266sp	19	3	3	3	6	6	4	3	St. Petersburg	Cervicitis
G/271sp	19	“	“	“	“	“	“	“	St. Petersburg	Cervicitis
G/273sp	19	“	“	“	“	“	“	“	St. Petersburg	Cervicitis
G/284sp	19	“	“	“	“	“	“	“	St. Petersburg	Cervicitis
G/286sp	19	“	“	“	“	“	“	“	St. Petersburg	Cervicitis
G/295sp	19	“	“	“	“	“	“	“	St. Petersburg	Cervicitis
G/301sp	19	“	“	“	“	“	“	“	St. Petersburg	Cervicitis
G/303sp	19	“	“	“	“	“	“	“	St. Petersburg	Cervicitis
G/315sp	19	“	“	“	“	“	“	“	St. Petersburg	Cervicitis
G/317sp	19	“	“	“	“	“	“	“	St. Petersburg	Cervicitis
G/225nl	19	“	“	“	“	“	“	“	The Netherlands	Cervix
G/227nl	19	“	“	“	“	“	“	“	The Netherlands	Cervix
G/229nl	19	“	“	“	“	“	“	“	The Netherlands	Cervicitis
G/230nl	19	“	“	“	“	“	“	“	The Netherlands	PID
G/SotonG1	19	“	“	“	“	“	“	“	United Kingdom	Cervicitis
B/193nl	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervicitis**
B/194nl	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**PID**
D/83s	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**San Francisco**	**Cervicitis**
D/202nl	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervix**
D/203nl	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervix**
D/204nl	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervix**
D/205nl	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervicitis**
D/206nl	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**PID**
D/207nl	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**PID**
D/SotonD5	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**United Kingdom**	**Cervicitis**
D/SotonD6	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**United Kingdom**	**Cervicitis**
E/279sp	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**St. Petersburg**	**Cervicitis**
E/294sp	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**St. Petersburg**	**Cervicitis**
H/114i	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**Indianapolis**	**Urethra**
H/115i	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**Indianapolis**	**Cervix**
H/18s	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**San Francisco**	**Cervicitis/urethritis**
H/40nl	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervicitis**
I/22p	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**Lisbon**	**Cervicitis/urethritis**
I/240nl	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**PID**
I/241nl	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervix**
I/245nl	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**PID**
J/44nl	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervicitis**
K/267sp	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**St. Petersburg**	**Cervicitis**
K/248nl	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervix**
K/249nl	**19**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervix**
D/43nl	**20**	**3**	**3**	**3**	**6**	**6**	**4**	**6**	**The Netherlands**	**Cervicitis and vaginal discharge**
G/13s	21	3	3	3	6	6	4	8	San Francisco	Proctitis
G/14s	21	“	“	“	“	“	“	“	San Francisco	Proctitis
A/Sa1	22	3	3	3	6	6	5	2	Saudi Arabia	Trachoma
Ia/57e	23	3	3	3	6	6	8	3	Ecuador	Cervicitis
Ia/94i	23	“	“	“	“	“	“	“	Indianapolis	Urethra

Ia/95i	23	“	“	“	“	“	“	“	Indianapolis	Cervix
Ia/118i	23	“	“	“	“	“	“	“	Indianapolis	Cervix
Ia/119i	23	“	“	“	“	“	“	“	Indianapolis	Urethra
Ia4/177i	23	“	“	“	“	“	“	“	Indianapolis	Urethra
Ia4/180i	23	“	“	“	“	“	“	“	Indianapolis	Cervix
Ia/178i	23	“	“	“	“	“	“	“	Indianapolis	Cervix
Ia/179i	23	“	“	“	“	“	“	“	Indianapolis	Urethra
Ia/183i	23	“	“	“	“	“	“	“	Indianapolis	Urethra
Ia/184i	23	“	“	“	“	“	“	“	Indianapolis	Cervix
Ia/SotonIa1	23	“	“	“	“	“	“	“	United Kingdom	Cervicitis
Ia/SotonIa3	23	“	“	“	“	“	“	“	United Kingdom	Cervicitis
Ia/UW202	23	“	“	“	“	“	“	“	Washington	Cervicitis
Ia/258b	23	“	“	“	“	“	“	“	Boston	
B/298sp	**23**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**St. Petersburg**	**Cervicitis**
D/2s	**23**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**San Francisco**	**Cervicitis/urethritis**
G/268sp	**23**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**St. Petersburg**	**Cervicitis**
H/231nl	**23**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervix**
H/232nl	**23**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervix**
H/233nl	**23**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervicitis**
H/234nl	**23**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**PID**
I/235nl	**23**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervix**
I/236nl	**23**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervix**
I/238nl	**23**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervicitis**
I/239nl	**23**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervicitis**
I/242nl	**23**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervix**
I/243nl	**23**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervicitis**
I/244nl	**23**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervicitis**
J/311sp	**23**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**St. Petersburg**	**Cervicitis**
J/247nl	**23**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervix**
K/251nl	**23**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervicitis**
Ia/24s	24	3	3	3	6	6	8	8	San Francisco	Cervicitis/urethritis
Ia/25s	24	“	“	“	“	“	“	“	San Francisco	Cervicitis/urethritis
D/UW3/Cx	25	3	3	3	6	7	4	1	Washington	Cervicitis
G/16p	**26**	**3**	**3**	**3**	**7**	**6**	**8**	**5**	**Lisbon**	**Cervicitis/urethritis**
G/17p	27	3	3	4	6	6	4	3	Lisbon	Cervicitis/urethritis
Ia/23p	28	3	3	4	6	6	4	4	Lisbon	Cervicitis/urethritis
H/21p	29	3	3	4	6	6	8	3	Lisbon	Cervicitis
G/UW57/Cx	30	3	3	5	6	6	4	3	Washington	Cervicitis
H/20p	31	3	3	7	6	6	8	3	Lisbon	Cervicitis/urethritis
F/38nl	32	4	4	3	2	7	4	3	The Netherlands	Cervicitis and vaginal discharge
L2b/48nl	33	5	2	3	8	1	4	11	The Netherlands	Proctitis
L2b/85nl	33	“	“	“	“	“	“	“	The Netherlands	Proctitis
F/ICCal3	34	6	3	3	2	7	4	3	Boston	
F/98i	34	“	“	“	“	“	“	“	Indianapolis	Urethra
F/99i	34	“	“	“	“	“	“	“	Indianapolis	Cervix
F/181i	34	“	“	“	“	“	“	“	Indianapolis	Urethra
F/182i	34	“	“	“	“	“	“	“	Indianapolis	Cervix
F/191i	34	“	“	“	“	“	“	“	Indianapolis	Urethra
F/192i	34	“	“	“	“	“	“	“	Indianapolis	Cervix
F/8p	34	“	“	“	“	“	“	“	Lisbon	Cervicitis/urethritis
F/9p	34	“	“	“	“	“	“	“	Lisbon	Cervicitis/urethritis
F/289sp	34	“	“	“	“	“	“	“	St. Petersburg	cervicitis
F/312sp	34	“	“	“	“	“	“	“	St. Petersburg	cervicitis
F/313sp	34	“	“	“	“	“	“	“	St. Petersburg	cervicitis
F/SW4	34	“	“	“	“	“	“	“	Sweden	Cervicitis
F/SW5	34	“	“	“	“	“	“	“	Sweden	Cervicitis
F/217nl	34	“	“	“	“	“	“	“	The Netherlands	Cervix
F/218nl	34	“	“	“	“	“	“	“	The Netherlands	Cervix
F/219nl	34	“	“	“	“	“	“	“	The Netherlands	Cervix

F/220nl	34	“	“	“	“	“	“	“	The Netherlands	Cervicitis
F/221nl	34	“	“	“	“	“	“	“	The Netherlands	PID
F/SotonF3	34	“	“	“	“	“	“	“	United Kingdom	Cervicitis
F/255b	34	“	“	“	“	“	“	“	Boston	
D/256b	**34**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**Boston**	**Cervicitis**
D2/96i	**34**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**Indianapolis**	**Urethra**
D2/97i	**34**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**Indianapolis**	**Cervix**
D2/189i	**34**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**Indianapolis**	**Cervix**
D2/190i	**34**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**Indianapolis**	**Urethra**
D/196nl	**34**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervix**
D/197nl	**34**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervix**
D/200nl	**34**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervicitis**
D/201nl	**34**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**PID**
D/209nl	**34**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervix**
D/SotonD1	**34**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**United Kingdom**	**Cervicitis**
E/19e	**34**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**Ecuador**	**Cervicitis**
E/5s	**34**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**San Francisco**	**Cervicitis/urethritis**
E/214nl	**34**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**PID**
Ja/41nl	**34**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervicitis and vaginal discharge**
Ja/47nl	**34**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervicitis and vaginal discharge**
J/252nl	**34**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**PID**
F/10s	35	6	3	3	2	7	4	8	San Francisco	PID
F/11s	35	“	“	“	“	“	“	“	San Francisco	PID
F/12s	35	“	“	“	“	“	“	“	San Francisco	PID
E/39nl	36	6	3	3	3	7	4	3	The Netherlands	Cervicitis
Da/TW448	**37**	**6**	**3**	**3**	**5**	**7**	**4**	**2**	**Taiwan**	**Trachoma**
D/3s	38	6	3	6	2	7	4	3	San Francisco	Cervicitis/urethritis
E/Bour	39	6	4	3	2	7	4	3	California	Cervicitis
E/28e	39	“	“	“	“	“	“	“	Ecuador	Cervicitis
E/55e	39	“	“	“	“	“	“	“	Ecuador	Cervicitis
E/56e	39	“	“	“	“	“	“	“	Ecuador	Cervicitis
E/88i	39	“	“	“	“	“	“	“	Indianapolis	Cervix
E/89i	39	“	“	“	“	“	“	“	Indianapolis	Urethra
E/102i	39	“	“	“	“	“	“	“	Indianapolis	Urethra
E/103i	39	“	“	“	“	“	“	“	Indianapolis	Cervix
E/106i	39	“	“	“	“	“	“	“	Indianapolis	Urethra
E/107i	39	“	“	“	“	“	“	“	Indianapolis	Cervix
E/108i	39	“	“	“	“	“	“	“	Indianapolis	Cervix
E/109i	39	“	“	“	“	“	“	“	Indianapolis	Urethra
E/110i	39	“	“	“	“	“	“	“	Indianapolis	Cervix
E/111i	39	“	“	“	“	“	“	“	Indianapolis	Urethra
E/116i	39	“	“	“	“	“	“	“	Indianapolis	Urethra
E/117i	39	“	“	“	“	“	“	“	Indianapolis	Urethra (female)
E6/120i	39	“	“	“	“	“	“	“	Indianapolis	Cervix
E6/121i	39	“	“	“	“	“	“	“	Indianapolis	Urethra
E/171i	39	“	“	“	“	“	“	“	Indianapolis	Urethra
E/172i	39	“	“	“	“	“	“	“	Indianapolis	Cervix
E/6p	39	“	“	“	“	“	“	“	Lisbon	Cervicitis
E/7p	39	“	“	“	“	“	“	“	Lisbon	Cervicitis
E/150	39	“	“	“	“	“	“	“	Seattle	Proctitis
E/11023	39	“	“	“	“	“	“	“	Seattle	Cervicitis
E/260b	39	“	“	“	“	“	“	“	Boston	
E/275sp	39	“	“	“	“	“	“	“	St. Petersburg	cervicitis
E/276sp	39	“	“	“	“	“	“	“	St. Petersburg	cervicitis
E/278sp	39	“	“	“	“	“	“	“	St. Petersburg	cervicitis
E/280sp	39	“	“	“	“	“	“	“	St. Petersburg	cervicitis
E/285sp	39	“	“	“	“	“	“	“	St. Petersburg	cervicitis
E/290sp	39	“	“	“	“	“	“	“	St. Petersburg	cervicitis

E/292sp	39	“	“	“	“	“	“	“	St. Petersburg	cervicitis
E/299sp	39	“	“	“	“	“	“	“	St. Petersburg	cervicitis
E/307sp	39	“	“	“	“	“	“	“	St. Petersburg	cervicitis
E/SW2	39	“	“	“	“	“	“	“	Sweden	Urethritis
E/SW3	39	“	“	“	“	“	“	“	Sweden	Cervicitis
E/58t	39	“	“	“	“	“	“	“	Tanzania	Conjunctivitis
E/45nl	39	“	“	“	“	“	“	“	The Netherlands	Cervicitis
E/213nl	39	“	“	“	“	“	“	“	The Netherlands	cervicitis
E/215nl	39	“	“	“	“	“	“	“	The Netherlands	PID
E/216nl	39	“	“	“	“	“	“	“	The Netherlands	PID
E/SotonE4	39	“	“	“	“	“	“	“	United Kingdom	Cervicitis
E/SotonE8	39	“	“	“	“	“	“	“	United Kingdom	Cervicitis
D/257b	**39**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**Boston**	**Cervicitis**
G/269sp	**39**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**St. Petersburg**	**Cervicitis**
E/4s	40	6	4	3	3	7	4	3	San Francisco	Cervicitis/urethritis
Ja/26s	**41**	**6**	**4**	**3**	**5**	**6**	**4**	**8**	**San Francisco**	**Cervicitis/urethritis**
C/31n	42	7	3	3	1	3	5	7	Nepal	Trachoma, TI
C/35n	43	7	3	3	6	3	5	7	Nepal	Trachoma, TI
C/1n	44	7	3	3	6	3	6	7	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
F/1-93	45	3	3	3	2	7	4	3	Seattle	Cervicitis
F/6-94	45	“	“	“	“	“	“	“	Seattle	Cervicitis
F/11-96	45	“	“	“	“	“	“	“	Seattle	Cervicitis
D1/90i	**45**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**Indianapolis**	**Cervicitis**
D1/91i	**45**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**Indianapolis**	**Urethra**
D/13-96	**45**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**Seattle**	**Cervix**
D/199nl	**45**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervicitis**
E/306sp	**45**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**St. Petersburg**	**Cervicitis**
E/92i	46	6	3	3	5	7	4	8	Indianapolis	Urethra
E/93i	46	“	“	“	“	“	“	“	Indianapolis	Cervix
E/104i	46	“	“	“	“	“	“	“	Indianapolis	Urethra
E/105i	46	“	“	“	“	“	“	“	Indianapolis	Cervix
E/173i	46	“	“	“	“	“	“	“	Indianapolis	Cervix
E/174i	46	“	“	“	“	“	“	“	Indianapolis	Urethra
E/185i	46	“	“	“	“	“	“	“	Indianapolis	Urethra
E/188i	46	“	“	“	“	“	“	“	Indianapolis	Cervix
D/210nl	**46**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**The Netherlands**	**Cervicitis**
E/100i	47	6	4	3	5	6	4	3	Indianapolis	Cervix
E/101i	47	“	“	“	“	“	“	“	Indianapolis	Urethra
D/EC	48	3	3	3	6	6	4	1	Montana	
D/LC	48	“	“	“	“	“	“	“	Montana	
K/SotonK1	**49**	**3**	**3**	**3**	**6**	**6**	**4**	**13**	**United Kingdom**	**Cervicitis**
B/Jali20	**50**	**3**	**3**	**3**	**6**	**6**	**5**	**7**	**Gambia**	**Trachoma**
G/11074	51	3	3	3	6	6	9	3	Seattle	Proctitis
G/9301	51	“	“	“	“	“	“	“	Seattle	Urethritis
G/9768	51	“	“	“	“	“	“	“	Seattle	Proctitis
G/11222	52	3	3	5	6	6	1	3	Seattle	Cervicitis
G/272sp	52	“	“	“	“	“	“	“	St. Petersburg	Cervicitis
G/223nl	52	“	“	“	“	“	“	“	The Netherlands	Cervix
G/228nl	52	“	“	“	“	“	“	“	The Netherlands	Cervicitis
A/HAR13	**53**	**3**	**5**	**3**	**6**	**3**	**5**	**7**	**Tunisia**	**Trachoma**
F/1	54	6	3	3	2	6	4	3	Seattle	Cervicitis
F/316sp	54	“	“	“	“	“	“	“	St. Petersburg	Cervicitis
F4/175i	55	6	3	3	2	7	4	12	Indianapolis	Cervix
F4/176i	55	“	“	“	“	“	“	“	Indianapolis	Urethra
J/151s	56	6	3	3	5	7	4	3	San Francisco	Cervicitis

L2/208nl	57	1	2	3	8	1	4	11	The Netherlands	PID
L2/246nl	58	3	3	3	8	1	4	11	The Netherlands	PID
G/222nl	59	3	6	5	6	6	1	3	The Netherlands	Cervix
G/224nl	59	“	“	“	“	“	“	“	The Netherlands	Cervix
D/195nl	**60**	**6**	**3**	**3**	**9**	**7**	**10**	**8**	**The Netherlands**	**Cervix**
E/212nl	61	6	4	3	2	9	4	3	The Netherlands	Cervix
D/198nl	**62**	**6**	**4**	**3**	**5**	**8**	**4**	**8**	**The Netherlands**	**Cervicitis**
G/226nl	**63**	**8**	**3**	**3**	**6**	**6**	**8**	**3**	**The Netherlands**	**Cervix**
E/211nl	64	9	4	3	2	7	4	3	The Netherlands	Cervix
K/291sp	**65**	**3**	**3**	**3**	**2**	**6**	**8**	**3**	**St. Petersburg**	**Cervicitis**
B/277sp	**66**	**3**	**3**	**3**	**2**	**6**	**12**	**18**	**St. Petersburg**	**Cervicitis**
G/270sp	67	3	3	3	6	6	4	15	St. Petersburg	Cervicitis
B/300sp	**68**	**3**	**3**	**3**	**6**	**6**	**8**	**6**	**St. Petersburg**	**Cervicitis**
G/302sp	**68**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**St. Petersburg**	**Cervicitis**
J/283sp	**69**	**3**	**3**	**3**	**6**	**6**	**8**	**14**	**St. Petersburg**	**Cervicitis**
Ia/262sp	70	3	3	3	6	6	8	16	St. Petersburg	Cervicitis
E/314sp	**71**	**3**	**3**	**3**	**6**	**6**	**11**	**3**	**St. Petersburg**	**Cervicitis**
D/263sp	**72**	**3**	**3**	**3**	**9**	**7**	**4**	**17**	**St. Petersburg**	**Cervicitis**
J/259b	**73**	**3**	**3**	**3**	**10**	**6**	**1**	**6**	**Boston**	**Cervicitis**
G/308sp	**74**	**3**	**3**	**3**	**11**	**7**	**4**	**3**	**St. Petersburg**	**Cervicitis**
E/264sp	75	3	4	3	2	7	4	3	St. Petersburg	Cervicitis
E/261sp	76	3	4	3	2	10	4	3	St. Petersburg	Cervicitis
E/265sp	76	**“**	**“**	**“**	**“**	**“**	**“**	**“**	St. Petersburg	Cervicitis
G/304sp	77	3	4	5	6	6	1	3	St. Petersburg	Cervicitis
D/297sp	**78**	**3**	**7**	**3**	**6**	**6**	**4**	**3**	**St. Petersburg**	**Cervicitis**
D/288sp	**79**	**6**	**3**	**3**	**9**	**7**	**4**	**8**	**St. Petersburg**	**Cervicitis**
D/310sp	**79**	**“**	**“**	**“**	**“**	**“**	**“**	**“**	**St. Petersburg**	**Cervicitis**
E/274sp	80	6	4	3	2	10	4	3	St. Petersburg	Cervicitis
E/287sp	80	“	“	“	“	“	“	“	St. Petersburg	Cervicitis
E/309sp	80	“	“	“	“	“	“	“	St. Petersburg	Cervicitis
E/281sp	81	6	4	8	2	7	4	3	St. Petersburg	Cervicitis
E/282sp	82	10	4	3	2	7	4	3	St. Petersburg	Cervicitis
E/293sp	83	11	4	3	2	7	4	3	St. Petersburg	Cervicitis
K/254b	**84**	**12**	**3**	**3**	**2**	**7**	**4**	**3**	**Boston**	**Cervicitis**

Note: **Boldface** denotes putative recombinant

† The strain ID is comprised of the letter that denotes the samples *omp*A genotype (e.g., E) followed by a backslash and then the number of the same that was assigned based on when it was added to the database (e.g., the first samples added have numbers 1, 2, 3, etc while samples added later have higher numbers. The last 1 or 2 letters denote the geographic region from which the sample derived and correlates with the region of isolation (e.g., sp, St. Petersberg; nl, The Netherlands; b, Boston).

Supplementary Table 4. Distribution of STs by geographic region.

ST	Africa n=21	Europe n=109	Russia n=58	Asia n=13	Americas n=122
1	3	13			9
2					1
3					1
4	1				
5		1			
6				1	
7					1
8					1
9					2
10	1				
11				1	
12	12				
13				2	
14					1
15		3	3		6
16					1
17					1
18					1
19		23	13		2
20		1			
21					2
22	1				
23		15	3		14
24					2
25					1
26		1			
27		1			
28		1			
29		1			
30					1
31		1			
32		1			
33		2			
34		20	3		15
35					3
36		1			
37				1	
38					1
39	1	10	10		24
40					1
41					1
42				1	

43				1	
44				6	
45		1	1		6
46		1			8
47					2
48					2
49		1			
50	1				
51					3
52		2	1		1
53	1				
54			1		1
55					2
56					1
57		1			
58		1			
59		2			
60		1			
61		1			
62		1			
63		1			
64		1			
65			1		
66			1		
67			1		
68			2		
69			1		
70			1		
71			1		
72			1		
73					1
74			1		
75			1		
76			2		
77			1		
78			1		
79			2		
80			3		
81			1		
82			1		
83			1		
84					1

Supplementary Table 5: eBURST report for 323 isolates*

Group 1: No. Isolates = 323 | No. STs = 84 | Predicted Founder = 19

ST	FREQ	SLV	DLV	TLV	SAT	Average Distance
19	41	15	24	17	27	2.75
23	32	14	18	15	36	3.2
39	46	10	10	6	57	4.21
34	38	9	13	10	51	3.65
21	2	8	20	23	32	3.13
20	1	7	22	18	36	3.2
75	1	7	7	15	54	3.81
48	2	6	22	19	36	3.22
24	2	6	20	15	42	3.57
68	2	6	20	15	42	3.65
49	1	5	23	19	36	3.24
67	1	5	23	19	36	3.24
32	1	5	8	8	62	4.43
83	1	5	8	8	62	4.43
82	1	5	8	8	62	4.43
64	1	5	8	8	62	4.43
11	1	5	5	23	50	4
51	3	4	27	16	36	3.33
17	1	4	27	16	36	3.33
71	1	4	27	16	36	3.33
70	1	4	21	16	42	3.68
69	1	4	21	16	42	3.68
45	8	4	18	26	35	3.25
15	13	4	16	23	40	3.72
27	1	4	16	21	42	3.59
29	1	4	14	17	48	4.03
56	1	4	11	17	51	3.8
35	3	4	9	16	54	4.02
46	9	4	6	12	61	4.18
8	1	3	19	28	33	3.19
30	1	3	19	19	42	3.57
16	1	3	18	22	40	3.65
31	1	3	14	18	48	4.06
5	1	3	14	18	48	4.06
52	4	3	11	21	48	4.09
36	1	3	11	16	53	3.85
80	3	3	10	9	61	4.46
79	2	3	6	12	62	4.21
43	1	3	4	3	73	4.59
7	1	2	21	27	33	3.15

50	1	2	21	23	37	3.66
9	2	2	14	25	42	3.63
3	1	2	13	29	39	3.48
78	1	2	13	29	39	3.49
84	1	2	13	15	53	3.86
40	1	2	11	13	57	4.42
61	1	2	11	9	61	4.48
55	2	2	10	13	58	4.13
12	12	2	8	23	50	4.03
76	2	2	7	13	61	4.07
13	2	2	5	26	50	4.07
37	1	2	5	13	63	4.27
44	6	2	2	4	75	4.66
59	2	2	1	13	67	4.83
77	1	2	1	13	67	4.66
41	1	2	1	10	70	4.56
57	1	2	1	0	80	5.42
22	1	1	19	24	39	3.72
65	1	1	17	23	42	3.51
10	1	1	14	27	41	3.54
54	2	1	13	31	38	3.46
63	1	1	13	20	49	3.81
25	1	1	12	34	36	3.4
14	1	1	11	28	43	3.57
74	1	1	10	33	39	3.46
81	1	1	10	9	63	5.07
38	1	1	9	12	61	4.5
28	1	1	8	23	51	4.07
47	2	1	7	14	61	4.19
53	1	1	4	5	73	4.73
73	1	1	3	19	60	4.25
62	1	1	2	8	72	5.01
60	1	1	2	6	74	4.79
42	1	1	2	5	75	5.12
1	25	1	2	0	80	5.42
33	2	1	2	0	80	5.43
18	1	0	20	22	41	3.78
26	1	0	6	21	56	4.21
72	1	0	4	21	58	3.92
58	1	0	4	13	66	4.09
2	1	0	3	25	55	3.81
6	1	0	2	7	74	4.6
4	1	0	2	3	78	4.97
66	1	0	1	23	59	4.13

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

Supplementary Table 6: Evidence of recombination for samples from Amsterdam, The Netherlands.

Sample	*ompA* genotype	ST sequence homology*	ST (majority *ompA*)	Location of SNPs	
L2/237nl	L2	L2	1 (L)		
K/250nl	L2	G or K	15 (K)	G: *pdhA,* 339; *pykF*, 40; *lysS*, 34; *leuS*: 282	K: *pdhA,* 339; *pykF*, 40; *lysS*, 34; *leuS*: 282
B/193nl	**B**	**G or K**	**19 (G)**	**G: *pdhA,* 339; *pykF*, 40**	**K: *pdhA,* 339; *pykF*, 40**
B/194nl	**B**	**G or K**	**19 (G)**	**G: *pdhA,* 339; *pykF*, 40**	**K: *pdhA,* 339; *pykF*, 40**
D/202nl	**D**	**G or K**	**19 (G)**	**G: *pdhA,* 339; *pykF*, 40**	**K: *pdhA,* 339; *pykF*, 40**
D/203nl	**D**	**G or K**	**19 (G)**	**G: *pdhA,* 339; *pykF*, 40**	**K: *pdhA,* 339; *pykF*, 40**
D/204nl	**D**	**G or K**	**19 (G)**	**G: *pdhA,* 339; *pykF*, 40**	**K: *pdhA,* 339; *pykF*, 40**
D/205nl	**D**	**G or K**	**19 (G)**	**G: *pdhA,* 339; *pykF*, 40**	**K: *pdhA,* 339; *pykF*, 40**
D/206nl	**D**	**G or K**	**19 (G)**	**G: *pdhA,* 339; *pykF*, 40**	**K: *pdhA,* 339; *pykF*, 40**
D/207nl	**D**	**G or K**	**19 (G)**	**G: *pdhA,* 339; *pykF*, 40**	**K: *pdhA,* 339; *pykF*, 40**
G/225nl	G	G or K	19 (G)	G: *pdhA,* 339; *pykF*, 40	K: *pdhA,* 339; *pykF*, 40
G/227nl	G	G or K	19 (G)	G: *pdhA,* 339; *pykF*, 40	K: *pdhA,* 339; *pykF*, 40
G/229nl	G	G or K	19 (G)	G: *pdhA,* 339; *pykF*, 40	K: *pdhA,* 339; *pykF*, 40
G/230nl	G	G or K	19 (G)	G: *pdhA,* 339; *pykF*, 40	K: *pdhA,* 339; *pykF*, 40
I/240nl	**I**	**G or K**	**19 (G)**	**G: *pdhA,* 339; *pykF*, 40**	**K: *pdhA,* 339; *pykF*, 40**
I/241nl	**I**	**G or K**	**19 (G)**	**G: *pdhA,* 339; *pykF*, 40**	**K: *pdhA,* 339; *pykF*, 40**
I/245nl	**I**	**G or K**	**19 (G)**	**G: *pdhA,* 339; *pykF*, 40**	**K: *pdhA,* 339; *pykF*, 40**
K/248nl	**K**	**G or K**	**19 (G)**	**G: *pdhA,* 339; *pykF*, 40**	**K: *pdhA,* 339; *pykF*, 40**
K/249nl	**K**	**G or K**	**19 (G)**	**G: *pdhA,* 339; *pykF*, 40**	**K: *pdhA,* 339; *pykF*, 40**
H/231nl	**H**	**Ia**	**23 (Ia)**		
H/232nl	**H**	**Ia**	**23 (Ia)**		

H/233nl	**H**	**Ia**	**23 (Ia)**		
H/234nl	**H**	**Ia**	**23 (Ia)**		
I/235nl	**I**	**Ia**	**23 (Ia)**		
I/236nl	**I**	**Ia**	**23 (Ia)**		
I/238nl	**I**	**Ia**	**23 (Ia)**		
I/239nl	**I**	**Ia**	**23 (Ia)**		
I/242nl	**I**	**Ia**	**23 (Ia)**		
I/243nl	**I**	**Ia**	**23 (Ia)**		
I/244nl	**I**	**Ia**	**23 (Ia)**		
J/247nl	**J**	**Ia**	**23 (Ia)**		
K/251nl	**K**	**Ia**	**23 (Ia)**		
D/196nl	**D**	**F**	**34 (F)**		
D/197nl	**D**	**F**	**34 (F)**		
D/200nl	**D**	**F**	**34 (F)**		
D/201nl	**D**	**F**	**34 (F)**		
D/209nl	**D**	**F**	**34 (F)**		
E/214nl	**E**	**F**	**34 (F)**		
J/252nl	**J**	**F**	**34 (F)**		
F/217nl	F	F	34 (F)		
F/218nl	F	F	34 (F)		
F/219nl	F	F	34 (F)		
F/220nl	F	F	34 (F)		
F/221nl	F	F	34 (F)		
E/213nl	E	E	39 (E)		
E/215nl	E	E	39 (E)		
E/216nl	E	E	39 (E)		
D/199nl	**D**	**F**	**45 (F)**	**F: *glyA*, 176; *glyA*, 264; *glyA*, 420**	
D/210nl	**D**	**E**	**46 (E)**		
G/223nl	G	G	52 (G)		
G/228nl	G	G	52 (G)		
L2/208nl	L2	L2	57 (L2)	L1, L2, L2a or L3: *mdhC*; 76	
L2/246nl	L2	L2	58 (L2)	L1, L2, L2a or L3: *glyA;* 215; *glyA*: 420; *mdhC*: 76; *mdhC*: 211	
G/222nl	G	G	59 (G)	G: *mdhC:* 194; *lysS:* 34	
G/224nl	G	G	59 (G)	G: *mdhC:* 194; *lysS:* 34	
D/195nl	**D**	**Da or F**	**60 (Da)**	**Da: *yhbG*: 267, 369, 451; *pykF*: 317; *lysS*: 280; *leuS*: 58; *leuS*: 58, 96**	**F: *yhbG*: 267, 369, 451; *pykF*: 317; *lysS*: 280; *leuS*: 58; *leuS*: 58, 96**

E/212nl	E	E	61 (E)	E: *pykF*: 484	
D/198nl	**D**	**Da or E**	**62 (Da)**	**Da: *yhbG*: 267, 451; *mdhC*: 183; *pykF*: 518, 521; *leuS*: 58; *leuS*: 96**	**E: *yhbG*: 267, 451; *mdhC*: 183; *pykF*: 518, 521; *leuS*: 58; *leuS*: 96**
G/226nl	**G**	**G or Ia**	**63 (G)**	**Ia: *glyA*: 191**	
E/211nl	E	E	64 (E)	E: *glyA*: 313	

Note: **Boldface** denotes putative recombinant; Occasionally both the *omp*A genotype and MLST ST match but where there are one or a few SNPs in the 7 genes.

**The seven ST genes of the sample have the highest homology to the seven genes of the strain that has the *omp*A genotype denoted in the column (e.g., for sample H/231nl that has a H *omp*A genotype, the seven ST genes were identical to the seven genes of Ia strains in the database where the *omp*A genotype was also Ia for those strains).

Supplemental Table 7: Evidence of recombination for samples from St. Petersburg, Russia.

Sample	*ompA* genotype	ST sequence homology	ST (majority *ompA*)	Location of SNPs	
E/261sp	E	E	76 (E)		
Ia/262sp	Ia	Ia	70 (Ia)		
D/263sp	**D**	**Da or F**	**72 (Da)**	**Da: *glyA:* 176 A→G, 264 C→T, 420 G→A *ybhG*: 451 A→G *leusS*: 58 G→A, 96 G→C, 422 G→A**	**F: *glyA:* 176 A→G, 264 C→T, 420 G→A *ybhG:* 267 G→A *leusS*: 58 G→A, 96 G→C, 422 G→A**
E/264sp	E	E	75 (E)		
E/265sp	E	E	76 (E)		
G/266sp	G	G or K	19 (G)	G: *pdhA*: 339 C→T	K: *pykF*: 40 G→A
K/267sp	**K**	**G**	**19 (K)**		
G/268sp	**G**	**Ia**	**23 (Ia)**		
G/269sp	**G**	**E**	**39 (E)**		
G/270sp	G	G or K	67 (G)	G: *pdhA*: 339 C→T *leusS*: 397 A→G	K: *pykF*: 40 G→A *leusS*: 397 A→G
G/271sp	G	G or K	19 (G)	G: *pdhA*: 339 C→T	K: *pykF*: 40 G→A
G/272sp	G	G	52 (G)		
G/273sp	G	G	19 (G)		
E/274sp	E	E	80 (E)		
E/275sp	E	E	39 (E)		
E/276sp	E	E	39 (E)		
B/277sp	**B**	**F**	**66 (F)**		
E/278sp	E	E	39 (E)		
E/279sp	**E**	**G or K**	**19 (G)**	**G: *pdhA*: 339 C→T**	**K: *pykF*: 40 G→A**
E/280sp	E	E	39 (E)		
E/281sp	E	E	81 (E)		
E/282sp	E	E	82 (E)		
J/283sp	**J**	**Ia**	**69 (Ia)**		
G/284sp	G	G	19 (G)		
E/285sp	E	E	39 (E)		
G/286sp	G	G or K	19 (G)	G: *pdhA*: 1 C→G, 339 C→T	K: *pdhA*: 1 C→G *pykF*: 40 G→A
E/287sp	E	E	80 (E)		
D/288sp	**D**	**Da or F**	**79 (Da)**	**Da: *ybhG*: 451 A→G *leusS*: 58 G→A, 96 G→C**	**F: *ybhG*: 267 G→A, 369 A→G *leusS*: 58 G→A**
F/289sp	F	F	34 (F)		
E/290sp	E	E	39 (E)		
K/291sp	**K**	**F or Ia**	**65 (F/Ia)**	**F: *glyA*: 176 A→G,**	**Ia: *ybhG*: 267 A→G,**

				264 C→T, 420 G→A ***pykF*: 317 C→T, 384 G→A** ***lysS*: 222 G→A, 261 T→C**	**303 A→G, 367 T→C, 369 G→A, 372 T→C, 387 T→C, 451 A→G**
E/292sp	E	E	39 (E)		
E/293sp	E	E	83 (E)		
E/294sp	**E**	**G**	**19 (G)**		
G/295sp	G	G	19 (G)		
K/296sp	K	K	15 (K)		
D/297sp	**D**	**G or K**	**78 (G/K)**	**G: *mdhC*: 168 G→A, 183 G→A, 339 C→T** ***lysS*: 82 A→T**	**K: *mdhC*: 168 G→A, 183 G→A** ***pdhA*: 339 C→T** ***lysS*: 82 A→T**
B/298	**B**	**Ia**	**23 (Ia)**		
E/299sp	E	E	39 (E)		
B/300sp	**B**	**Ia**	**68 (Ia)**		
G/301sp	G	G	19 (G)		
G/302sp	**G**	**A or Ba**	**68 (A/Ba)**	**A: *mdhC:* 168 G→A, 183 G→A** ***lysS:* 222 G→A** ***leuS:* 96 G→C, 282 T→C**	**Ba: *mdhC:* 168 G→A,** **183 G→A** ***lysS:* 202 G→A, 222 G→A** ***leuS:* 58 A→G**
G/303sp	G	G or I	19 (G)	G: *pdhA*: 339 C→T	I: *pykF*: 40 G→A
G/304sp	G	G	77 (G)		
K/305sp	K	G or K	15 (K)	G: *pdhA*: 339 C→T *lysS*: 34 A→G *leusS*: 282 T→C	K: *pykF*: 40 G→A *lysS*: 34 A→G *leusS*: 282 T→C
E/306sp	**E**	**F**	**45 (F)**		
E/307sp	E	E	39 (E)		
G/308sp	**G**	**D**	**74 (D)**		
E/309sp	E	E	80 (E)		
D/310sp	**D**	**Da or F**	**79 (Da)**	**Da: *ybhG*: 451 A→G** ***leusS*: 58 G→A, 96 G→C**	**F: *ybhG*: 267 G→A, 369A→G** ***leusS*: 58 G→A**
J/311sp	**J**	**Ia**	**23 (Ia)**		
F/312sp	F	F	34 (F)		
F/313sp	F	F	34 (F)		
E/314sp	**E**	**G or K**	**71 (G/K)**	**G: *pdhA*: 339 C→T** ***lysS*: 221 G→A, 470 G→A**	**K: *pykF*: 40 G→A** ***lysS*: 221 G→A, 470 G→A**
G/315sp	G	G	19 (G)		
F/316sp	F	F	54 (F)		

G/317sp	G	G or K	19 (G)	G: *pdhA*: 339 C→T *lysS*: 34 A→G	K: *pykF*: 40 G→A *lysS*: 34 A→G
J/318sp	**J**	**G or K**	**15 (G/K)**	**G: *pdhA*: 339 C→T** ***lysS*: 34 A→G**	**K: *pykF*: 40 G→A** ***lysS*: 34 A→G**

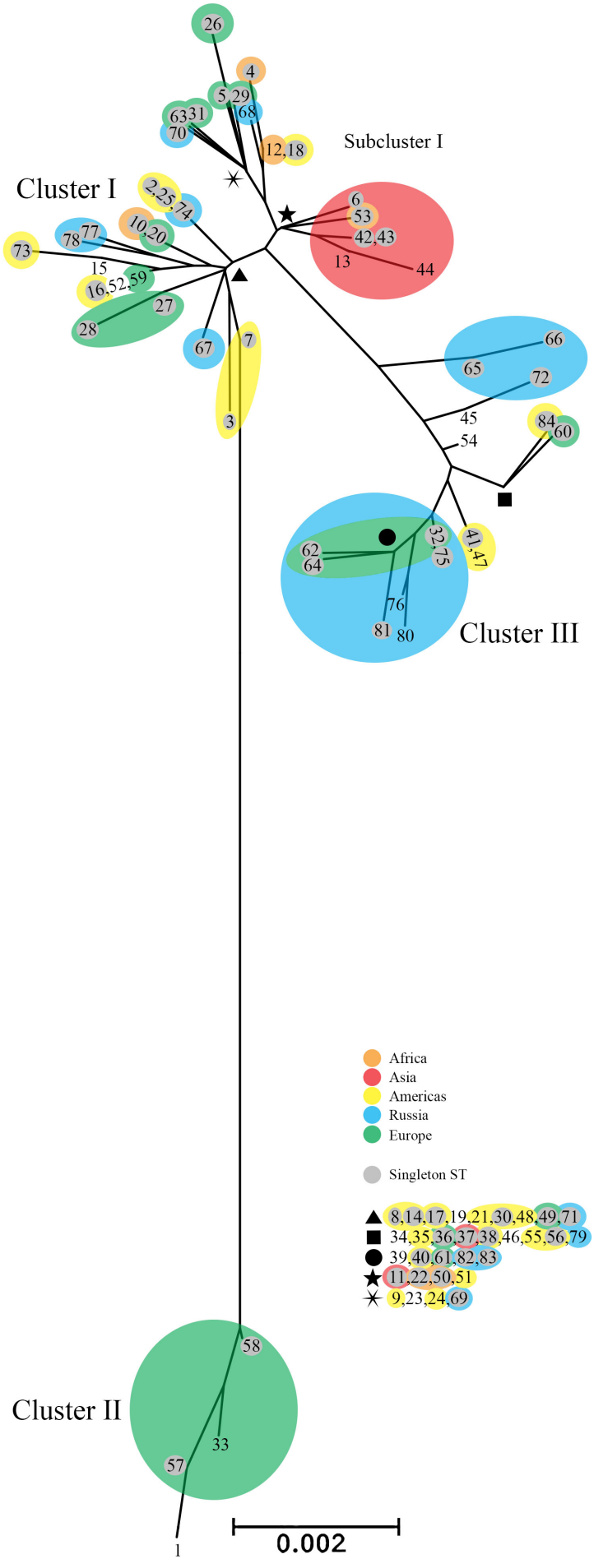
Note: **Boldface** denotes putative recombinant

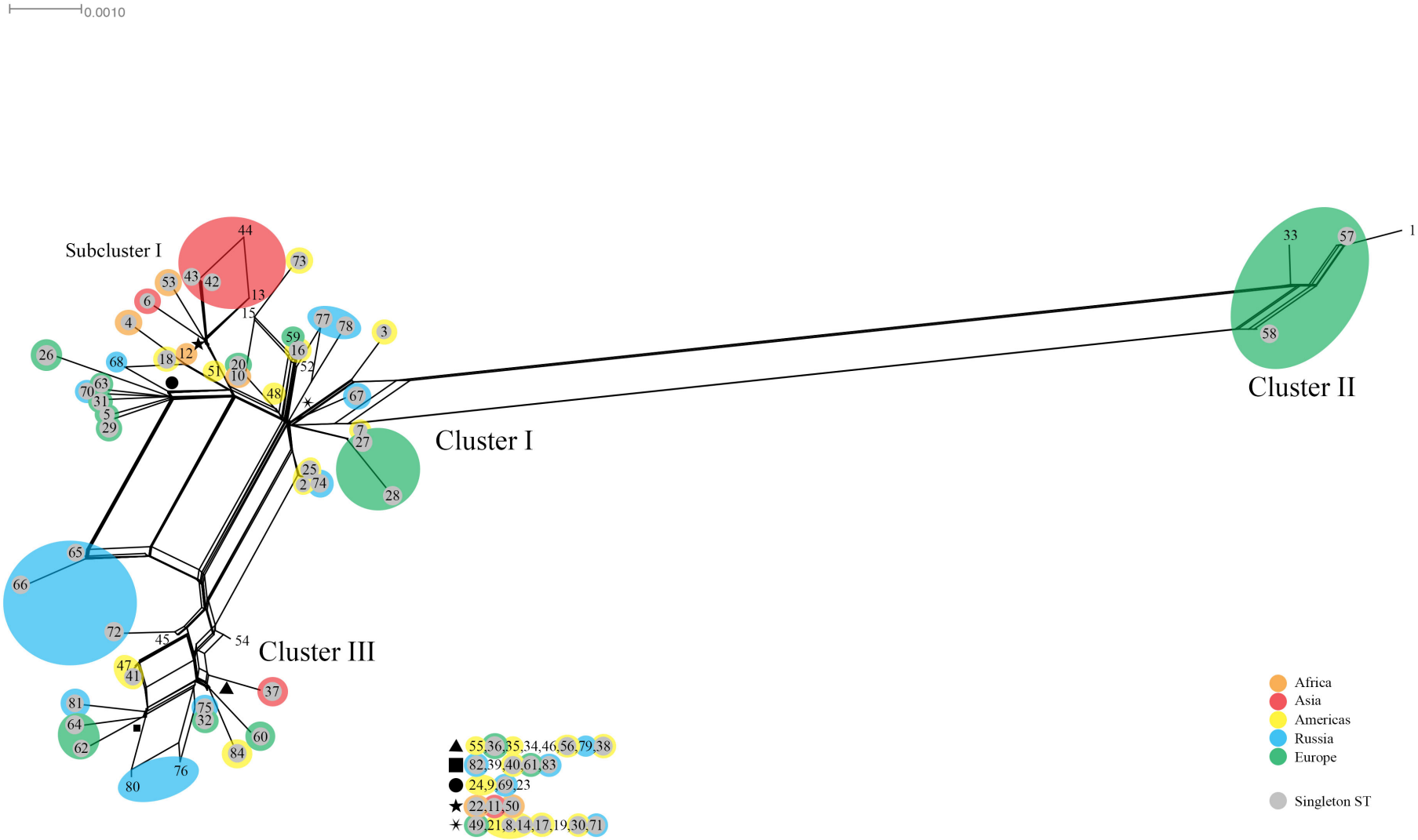
*The seven ST genes of the sample have the highest homology to the seven genes of the strain that has the *omp*A genotype denoted in the column (e.g., for sample G/268sp that has a G *omp*A genotype, the seven ST genes were identical to the seven genes of Ia strains in the database where the *omp*A genotype was also Ia for those strains).

Supplementary Table 8: Pairwise Population Differentiation (Fst) for North American, Dutch, and Russian Women

Comparison	Fst
American vs. Netherlands Women	0.05332*
Russian vs. American Women	0.00966 ns
Russian vs. Netherlands Women	0.01996 ns
Russian vs. Netherlands Women non-LGV	0.01101 ns
American vs Netherlands Women non-LGV L2	0.04671*

*$p < 0.05$ by permutation test (n = 110)


Cluster I
Subcluster I
Cluster III
Cluster II
26
4
5,29
68
63
31
70
12,18
6
53
42,43
13
44
2,25,74
10,20
78
77
73
15
16,52,59
27
28
67
7
3
66
65
72
45
54
84
60
62
64
32,75
41,47
76
81
80
58
33
57
1
0.002
Africa
Asia
Americas
Russia
Europe
Singleton ST
8,14,17,19,21,30,48,49,71
34,35,36,37,38,46,55,56,79
39,40,61,82,83
11,22,50,51
9,23,24,69


0.0010
Subcluster I
Cluster I
Cluster II
Cluster III
▲ 55,36,35,34,46,56,79,38
■ 82,39,40,61,83
● 24,9,69,23
★ 22,11,50
✶ 49,21,8,14,17,19,30,71
Africa
Asia
Americas
Russia
Europe
Singleton ST