Chlamydia trachomatis: Genome sequence analysis of lymphogranuloma venereum isolates

Nicholas R. Thomson,[1,7] Matthew T.G. Holden,[1] Caroline Carder,[2] Nicola Lennard,[1] Sarah J. Lockey,[3] Pete Marsh,[4] Paul Skipp,[5] C. David O'Connor,[5] Ian Goodhead,[1] Halina Norbertzcak,[1] Barbara Harris,[1] Doug Ormond,[1] Richard Rance,[1] Michael A. Quail,[1] Julian Parkhill,[1] Richard S. Stephens,[6] and Ian N. Clarke[4]

[1]The Pathogen Sequencing Unit, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK; [2]Chlamydia and Molecular Laboratory, Department of Clinical Microbiology, University College Hospitals NHS Foundation Trust, The Windeyer Institute of Medical Sciences, London, W1T 4JF, UK; [3]Molecular Microbiology Group, University Medical School, Southampton General Hospital, Southampton, SO16 6YD, UK; [4]Health Protection Agency South East, Southampton Laboratory, Southampton General Hostpital, Southampton, SO16 6YD, UK; [5]Centre for Proteomic Research, School of Biological Sciences, University of Southampton, Southampton, SO16 7PX, UK; [6]School of Public Health, University of California at Berkeley, Berkeley, California 94720, USA

Chlamydia trachomatis is the most common cause of sexually transmitted infections in the UK, a statistic that is also reflected globally. There are three biovariants of *C. trachomatis*: trachoma (serotypes A–C) and two sexually transmitted pathovars; serotypes D–K and lyphogranuloma venereum (LGV). Trachoma isolates and the sexually transmitted serotypes D–K are noninvasive, whereas the LGV strains are invasive, causing a disseminating infection of the local draining lymph nodes. Genome sequences are available for single isolates from the trachoma (serotype A) and sexually transmitted (serotype D) biotypes. We sequenced two isolates from the remaining biotype, LGV, a long-term laboratory passaged strain and the recent "epidemic" LGV isolate-causing proctitis. Although the genome of the LGV strain shows no additional genes that could account for the differences in disease outcome, we found evidence of functional gene loss and identified regions of heightened sequence variation that have previously been shown to be important sites for interstrain recombination. We have used new sequencing technologies to show that the recent clinical LGV isolate causing proctitis is unlikely to be a newly emerged strain but is most probably an old strain with relatively new clinical manifestations.

[Supplementary material is available online at www.genome.org. The genome sequence data from this study has been submitted to EMBL: AM884176, AM886278, AM884177, and AM886279.]

Chlamydia trachomatis is the major cause of sexually transmitted infections (STIs) globally with an estimated 89 million cases in 1995 (Peeling and Brunham 1996). *C. trachomatis* isolates are classified serologically with 15 serovariants, based on the major outer membrane protein (OmpA). Trachoma is caused by serovars A–C, whereas serovars D–K and L1, L2, and L3 are associated with sexually transmitted infections. Serovars D–K cause cervicitis in women and urogenital infections in men, and L1, L2, and L3 represent the three different serovars causing lymphogranuloma vernerum (LGV).

C. trachomatis-causing LGV are much more invasive than serovars A–K; they cause systemic infections, infect monocytes, and disseminate to the local lymph nodes, where they can cause large swellings characteristic of bubonic diseases. LGV serovars are endemic in parts of Africa, South East Asia, South America, and the Caribbean (Viravan et al. 1996; Behets et al. 1999; Mabey and Peeling 2002). The incidence of LGV in the West is low, and most of the reported cases are likely to have been imported following travel to endemic countries. However, a recent European "outbreak" of LGV in men who have sex with men (MSM) has been reported (Nieuwenhuis et al. 2003). These patients presented with proctitis rather than genital ulceration and the typical inguinal buboes, characteristic of LGV in the tropics, which were absent. These less obvious clinical features likely explain the initial difficulties encountered in describing the outbreak. It has been speculated that the European outbreak LGV strains causing proctitis represent a new emerging infection. This is supported by the observed variations in the *ompA* gene (*ompA* has been used to "genotype" chlamydial isolates), which indicates that these European proctitis isolates are a new variant of the LGV serovar L2, known as L2b. Conversely, it has also been suggested that rectal chlamydial infections are common, but their detection has not, until recently, become part of a routine and accredited testing protocol. This may infer, as has been suggested, that rectal LGV has been present for years and to attribute the newly observed proctitis infections to a new epidemic is incorrect (Schachter and Moncada 2005). In support of this, L2b was identifed in San Francisco in the 1980s (Spaargaren et al. 2005).

Complete genome sequences are only available for a single ocular and a single genital-tract isolate of *C. trachomatis*, Har-13 and UW-3, respectively (Stephens et al. 1998; Carlson et al. 2005). Comparison of these genomes revealed a very high level of

[7]Corresponding author.
E-mail nrt@sanger.ac.uk; fax 44-1223-494919.
Article published online before print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.7020108.

Table 1. General properties of *C. trachomatis* genomes

	C. trachomatis			
Strain	**UW-3[a]**	**Har-13[b]**	**L2**	**UCH-1**
Serovar	D	A	L2	L2b
Biovariant	Genital	Ocular	LGV	LGV
Size	1,042,519	1,044,459	1,038,842	1,038,869
% G+C content	41.27	41.27	41.33	41.33
N° CDSs	894	920	889	889
Coding density	90	90	89	89
Average gene size	1050	1032	1052	1052
rRNA operons	2	2	2	2
tRNA	37	37	37	37
Pseudogenes	5	8	15	15

(LGV) Lymphogranuloma venerium.
[a]Stephens et al. (1998).
[b]Carlson et al. (2005).

synteny and genome-sequence conservation, although some gene differences were identified that may account for the differing disease outcomes associated with the different biovars (Caldwell et al. 2003; Carlson et al. 2004). Thus, comparison of an LGV isolate with these related strains is likely to provide further insights into the mechanisms that may influence disease outcome and the underlying genetic differences that differentiate these biovariants.

The aim of this study was to complete the genomic analysis of the remaining *C. trachomatis* biovariant, LGV. The LGV isolate sequenced, strain L2/434/Bu (referred to as *C. trachomatis* strain L2) was originally isolated from an inguinal bubo of a LGV case 40 yr ago in California (Schachter and Moncada 2005). In addition, we also obtained a recent clinical LGV isolate from a patient with proctitis in London (Table 1). Comparison of the genomes of these two strains should indicate the evolutionary origins of these isolates and answer the question of whether the proctitis "outbreak" is due to a new emerging *C. trachomatis* LGV biovariant or part of an ongoing, but until recently, undetected disease.

Results

General features of the *C. trachomatis* strain L2 genome and comparisons with ocular and genital isolates

The genome of *C. trachomatis* strain L2 is composed of a single circular chromosome of 1.039 Mb and a plasmid of 7499 bp, predicted to encode 889 and eight coding sequences (CDS), respectively (Fig. 1; Table 1). The plasmid was identical in sequence and gene content to that previously published for this strain (Comanducci et al. 1988).

Comparison of the strain L2 genome sequence with those of strains *C. trachomatis* strain UW-3 (strain UW-3) and *C. trachomatis* strain Har-13 (strain Har-13) (Stephens et al. 1998; Carlson et al. 2005) showed that the genomes of all three isolates are similar in size, the number of CDSs, and nucleotide composition (Table 1). A detailed comparison of the predicted CDSs showed that 846 CDSs were apparently intact and common to all three genomes (Fig. 2). Moreover, the median nucleotide identity of CDS shared between strain L2 and either Har-13 or UW-3 was 99.53% and 99.55%, respectively. CDSs that were present in two or less of the three genomes could be entirely accounted for by either in silico differences in gene prediction, or as a consequence of functional gene loss (deletion and pseudogene formation) (Table 2), indicating that recent gene acquisition has not played a role in changes in disease causation in this group of organisms. To highlight possible phenotypic differences in the coding capacity of these *C. trachomatis* strains, pseudogenes were not included in our comparisons described by Figure 2, but they are described in detail in Supplemental Table 1 and discussed below.

Wider comparisons were also performed between strain L2 and more distant members of the *Chlamydiaceae*, including *Chlamydophila pneumoniae* (Read et al. 2000), *Chlamydia muridarum* (Read et al. 2000), *Chlamydophila caviae* (Read et al. 2003), and *Chlamydophila felis* (Azuma et al. 2006) (Fig. 1). These members of the *Chlamydiaceae* are associated with hosts as diverse as humans and frogs and can cause infections ranging from human adult-onset asthma to feline conjunctivitis (Saikku 1999; Longbottom and Coulter 2003). Despite the considerable variations in disease and host range, the chlamydial genomes are remarkably similar in content: 834, 784, 783, and 777 orthologs of strain L2 CDSs were detected in *C. muridarum*, *Cp. felis*, *Cp. Caviae*, and *Cp. pneumoniae*, respectively (Fig. 1).

These whole-genome comparisons agreed with previous findings showing that the *Chlamydiaceae* are extremely similar in gene content, even amongst more distantly related isolates (Read et al. 2003; Carlson et al. 2005; Thomson et al. 2005). Since there was no evidence of recent gene acquisition in strain L2, we concentrated our analysis on the apparent loss of gene function (Table 2).

Evidence of functional loss

The plasticity zone (PZ)

The plasticity zone (PZ) is the site of the most extensive variation in sequence and gene content between chlamydial genomes (Read et al. 2000, 2003). The variation in size of the *C. trachomatis* PZ is largely due to differential deletion of the cytotoxin gene(s), which have been almost entirely deleted from strain L2, leaving two gene remnants CTL0420 and CTL0421 (Belland et al. 2001; Carlson et al. 2004) (Fig. 3).

However, there is further evidence for genome decay in strain L2; CTL0426A is a remnant formed by the likely deletion of two CDS, orthologs of which (CT173/CTA_0189 and CT174/CTA_190; unknown function) remain present and intact in strains UW-3 and Har-13 (Fig. 3). Other CDSs within the PZ that show variation include those encoding Phospholipase D (PLD). The PZ encodes four PLD CDSs (CTL0409, CTL0411, CTL0413, and CTL0414; PLD locus 2), which show extensive variation in sequence between isolates. Whilst orthologs of CTL0411 and CTL0413 in *C. trachomatis* strains UW-3 and Har-13 appear intact and conserved in sequence, those for CTL0409 and CTL0414 have suffered multiple independently acquired frameshift mutations and deletion events (summarized in Fig. 3).

The PZ PLDs are known to be unique to *C. trachomatis* and the related rodent pathogen, *C. muridarum* (Nelson et al. 2006). However, there are two other non-PZ-located PLDs in strain L2: CTL0339 and CTL0536 located in locus 1 and 3, respectively (Fig. 1; Nelson et al. 2006; this study). The PLDs from locus 1 and 3 are

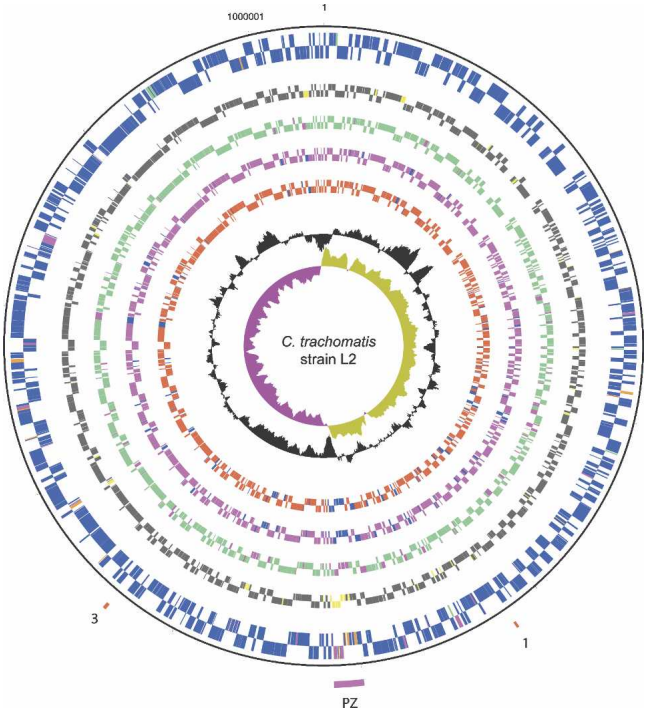

Figure 1. Circular representation of the *C. trachomatis* strain L2 chromosome. The outer scale shows the size in base pairs. From the outside in, circles 1 and 2 show the position of CDSs transcribed in a clockwise and anticlockwise direction. Using the published gene predictions for *C. trachomatis* strains UW-3 and Har-13, the strain L2 CDSs have been color-coded depending on whether they are: (blue) predicted and intact in all isolates; (pink) predicted and intact in L2 and UW-3; (green) predicted and intact in L2 and Har-13; (orange) defunct in L2, predicted and intact in Har-13 and UW-3; (red) unique to L2; (brown) defunct in all isolates. (Circles 3–10) *C. trachomatis* strain L2 CDSs that are present/absent in: *C. muridarum* (circles 3 and 4; present gray; absent yellow), *Cp. felis* (circles 5 and 6; present green; absent pink), *Cp. caviae* (circles 7 and 8; present pink; absent blue), and *Cp. pneumoniae* (circles 9 and 10; present red; absent blue) by reciprocal FASTA analysis. Circle 11 shows a plot of G+C content (in a 0.5-kb window); circle 12 shows a plot of GC skew ([G-C]/[G+C]; in a 0.5-kb window). The position of the PZ (pink) and the PLD CDSs in locus 1 and 3 are numbered accordingly and marked (red). Since the gene content of strain L2 and UCH-1 are essentially identical, these data apply equally to both isolates.

conserved in *Chlamydiaceae*, with orthologs present in most other sequenced chlamydial genomes. Phylogenetic analysis of all of the *C. trachomatis* PLDs showed that the PLDs from the different loci are phylogenetically distinct (Fig. 4). Interestingly, the topology of the phylogenetic tree for the PZ PLDs suggests that they have arisen by paralogous expansion since the divergence from a common ancestor.

Metabolism

There is evidence of functional gene loss outside of the PZ region of strain L2. CDSs CTL0228 and CTL0856 are pseudogenes of fumarate hydratase (*fumC*; fumarase C) and succinate dehydrogenase (*sdhC*; cytochrome B subunit), respectively. Fumarases are involved in the citric acid cycle, either reversibly converting fumarate into L-malate or converting oxaloacetate into succinate, depending on oxygen availability (Guest and Roberts 1983; Guest et al. 1985). Whilst *fumC* appears intact in strains UW-3 and Har-13 (CTA_0932 and CT855, respectively), the strain L2 ortholog, CTL0228, carries two frameshift mutations. Our findings further confirm the previous observations, in which a truncated FumC was detected in strain L2 during 2D gel/proteomic analysis (Shaw et al. 2002).

Like FumC, succinate dehydrogenase is also required for the metabolism of fumarate by facilitating the aerobic interconversion of fumarate and succinate. In *Escherichia coli*, succinate dehydrogenase is encoded by four genes, *sdhA–D*. *sdhA* and *sdhB* encode the succinate dehydrogenase flavoprotein and iron–sulphur protein subunits, respectively. SdhA and SdhB are both anchored to the cytoplasmic membrane by SdhC and SdhD. SdhC also possesses a cytochrome b556 domain involved in electron transport.

Analysis of the *C. trachomatis* genome sequences showed that all carried orthologs of the genes *sdhA*, *sdhB*, and *sdhC* (CTL0854–CTL0856), but none possessed *sdhD* homologs. However, it is evident that whilst the *sdhA* and *sdhB* remain intact, the orthologs of *sdhC* in all of the sequenced *C. trachomatis* carry the same frameshift mutation (following codon 103), as well as other additional point mutations (Table 2). This appears to be specific to *C. trachomatis*, since these genes are intact in the other sequenced *Chlamydiaceae* (data not shown).

Other potentially important pseudogenes lie within a cluster of two CDSs encoding a pyruvoyl-dependent arginine decarboxylase (CTL0627; PvlArgDC) and an arginine/ornithine antiporter (CTL0628/ArcD) (Graham et al. 2002). *PvlArgDC* and *arcD* are present in all of the three *C. trachomatis* genomes, and both CDSs are intact in strain UW-3 (CT373 and CT374, respectively). However, the CTL0627 (PvlArgDC) carries a mutation in strain L2 and the ortholog of *arcD* (encompassing CDS CTA_0406-CTA_0408) carries multiple mutations in strain Har-13.

The PvlArgDC family of enzymes are involved in the biosynthesis of putrescine from L-arginine. Since *Chlamydia* spp. lack any other genes for arginine catabolism (Graham et al. 2002), it has been suggested that these genes may be involved in uptake of host L-arginine, along with a proton. Arginine is predicted to be degraded to agmatine, which is subsequently exported from the cell. The process of uptake and degradation of arginine is thought to play a role in pH homeostasis, both in *Chlamydia* and other bacterial pathogens such as the salmonellae (Graham et al. 2002; Kieboom and Abee 2006). This effect is mediated by the uptake of protons and concomitant depletion of acidic products made from carbohydrate fermentation by the host.

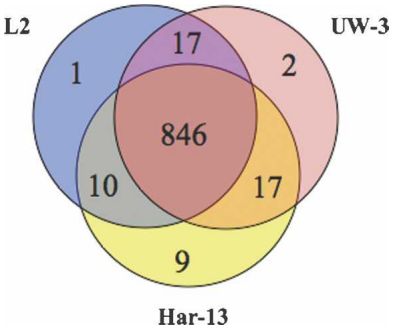

Figure 2. Distribution of predicted and functional CDSs shared between *C. trachomatis* strains L2, UW-3, and Har-13. The Venn diagram shows the number of CDS predicted to be functional, that are unique or shared, between one or more *C. trachomatis* isolates (see Methods). Pseudogenes were counted as absent in this analysis.

Table 2. A description of the *C. trachomatis* strain L2 pseudogenes identified by whole-genome comparisons with *C. trachomatis* strains UW-3 and Har-13

C. trachomatis strain L2 pseudogenes				
L2 CDS	Orthologs in other strains			Product description
	UCH-1	Har-13	UW-3	
CTL0161	CTLon_0162[a]	CTA_0863	CT793	Conserved hypothetical protein
CTL0228	CTLon_0228[a]	CTA_0932	CT855	Fumarate hydratase (FumC)
CTL0292	CTLon_0287[a]	CTA_0039	CT037	Conserved hypothetical protein
CTL0409	CTLon_0405[a]	CTA_0163	CT154	Phospholipase D protein
CTL0414	CTLon_0408[a]	CTA_0167	CT158	Phospholipase D protein
CTL0415	CTLon_0409[a]	CTA_0169	CT160	Conserved hypothetical protein
CTL0418	CTLon_0411[a]	CTA_0171	CT162	Putative membrane protein
CTL0420	CTLon_0414[a]	CTA_0179	CT167	Cytotoxin (adherence factor)
CTL0421	CTLon_0415[a]	CTA_0181	CT168	Cytotoxin (adherence factor)
CTL0426A	CTLon_0421[a]	CTA_0190	CT174	Conserved hypothetical protein
CTL0552	CTLon_0548[a]	CTA_0322	CT300	Putative integral membrane protein
CTL0578	CTLon_0574[a,b]	CTA_0350	CT326	Conserved hypothetical protein
CTL0612	CTLon_061[a]	CTA_0389	CT358	Inner membrane protein
CTL0627	CTLon_0625[a]	CTA_0405	CT373	Pyruvoyl-dependent arginine decarboxylase
CTL0856	CTLon_0850[a]	CTA_0643[c] CTA_0644[c]	CT593[c] CT593.1[c]	Succinate dehydrogenase (SdhC)

[a]Pseudogene.
[b]Lacks the first frameshift mutation carried by CTL0578, following codon 58.
[c]Originally annotated as separate CDS.

resequencing, lies within a homopolymer, or represents a low-confidence base call.

Analysis of the coding capacity of strain UCH-1 showed that it was identical to strain L2 (Table 1). This extended to the complement of pseudogenes; with only one exception, CTLon_0574 (CTLon_0574 carries 2/3 mutations found in its strain L2 pseudogene ortholog CTL0578); all of the strain UCH-1 pseudogenes exhibited the same defining mutation(s) carried by their defunct strain L2 orthologs (Table 2).

Based on these data, we estimate that there are 573 high-confidence base differences that distinguish strain UCH-1 from strain L2. Of these bases, 457 occur within CDSs, which is slightly fewer than would be expected assuming a random distribution (data not shown). The most notable sequence variation (because of its possible utility for diagnostics) is found within a CDS that encodes the translocated actin recruiting phosphoprotein (Tarp; CTLon_0712). *C. trachomatis* entry into the host cell is a result of chlamydial reconfiguration of the host cell's actin skeleton, and it is proposed that Tarp contributes to the pathogen-directed phagocytosis, i.e., the uptake of elementary bodies (EBs) (Jewett et al. 2006). The differences between Tarp in *C. trachomatis* strains L2, UW-3, and Har-13 (i.e., two large deletions and a high number of SNPs) has already been noted (Carlson et al. 2005), and may be significant in accounting for the difference in tropism between LGV, the trachoma (serovars A–C), and sexually transmitted serovars D–K. The Tarp sequences of strain UCH-1 and strain L2 are more similar to each other than the *C. trachomatis* serovars A–K Tarp sequences. However, there are a number of notable differences in sequence between Tarp from strain L2 and UCH-1: a deletion of 12 bases between nucleotides 718 and 731 of strain L2, resulting in the loss of the four amino acids: asparagine, isoleucine, tyrosine, and glutamic acid. There is a three-base in-frame insertion between bases 480 and 481 of strain L2, resulting in an extra alanine at this point, and a three-base in-frame deletion between bases 933 and 937, resulting in the loss of an alanine. The latter occur in one of the six repeat units in the 5′ insertion/deletion-coding region of Tarp (Carlson et al. 2005). Apart from the aforementioned differences, there are three SNPs between strains UCH-1 and L2, which are actually conserved in the corresponding nucleotide positions in *C. trachomatis* strains UW-3 and Har-13, suggesting a role for intra-serovar recombination in the diversity of these genes, as has been reported for the *Chlamydia* (Gomes et al. 2007). There is only one SNP in strain UCH-1 (from a "C" residue in *C. trachomatis* strains L2, UW-3, and Har-13 to "T" in strain UCH-1) that appears to be unique. This, in addition to the 12-base deletion mentioned earlier, are potential targets for differential molecular diagnostic analysis, for example, by use of FRET-probe based real-time PCR, for the discrimination of L1, L2, and L3 from the proctitis strain UCH-1.

These data show that when comparing the classical LGV isolate and the recent clinical isolate causing proctitis, there is no evidence of gene gain or obvious mutations that would indicate further functional loss. Therefore, the only differences between

Recent clinical *C. trachomatis* LGV isolate causing proctitis: Biological properties and sequence

The newly described epidemics of LGV proctitis afforded a unique opportunity to compare a recent clinical LGV strain, with atypical clinical manifestations, with that of the classical LGV strain L2. We isolated and sequenced an LGV strain from a patient with proctitis in London in 2006: strain L2/UCH-1/proctitis (referred to as strain UCH-1; see Methods). Comparisons of the growth rate and inclusion formation of strain UCH-1 with the classical LGV strain L2, passaged in tissue culture over many decades, showed no morphological differences or increase/decrease in replication proficiency (data not shown).

Strain UCH-1 was sequenced using a pyrosequencing approach (454/Roche GS20). In total, 691 single base-pair differences were detected using MUMmer (Delcher et al. 1999) in strain UCH-1 compared with strain L2. Of these, 123 were found within homopolymeric tracts and 568 were either base substitutions or indels. All of the bases that were predicted to cause premature termination or frameshifts in CDSs otherwise conserved between strain L2 and strain UCH-1 were resequenced using PCR amplification and capillary sequencing. In total, we resequenced 71 single nucleotide polymorphisms (SNPs) that lay within homopolymeric tracts, 66/71 of which were incorrect 454 base calls. Conversely, we resequenced 48 SNPs and/or indels, all of which were correctly called. This is consistent with substitution errors being relatively rare in 454/Roche pyrosequencing data, while errors in base calls extending or contracting the length of homopolymeric tracts being more frequent (Margulies et al. 2005). We also resequenced all of the mutations in all of the strain UCH-1 pseudogenes (see below).

However, since we did not resequence all of the remaining called differences between strains, the sequence should be considered to be a contiguous draft. A summary of all of the base-pair differences can be found in Supplemental Table 2, along with a comment on whether the base has been confirmed by capillary

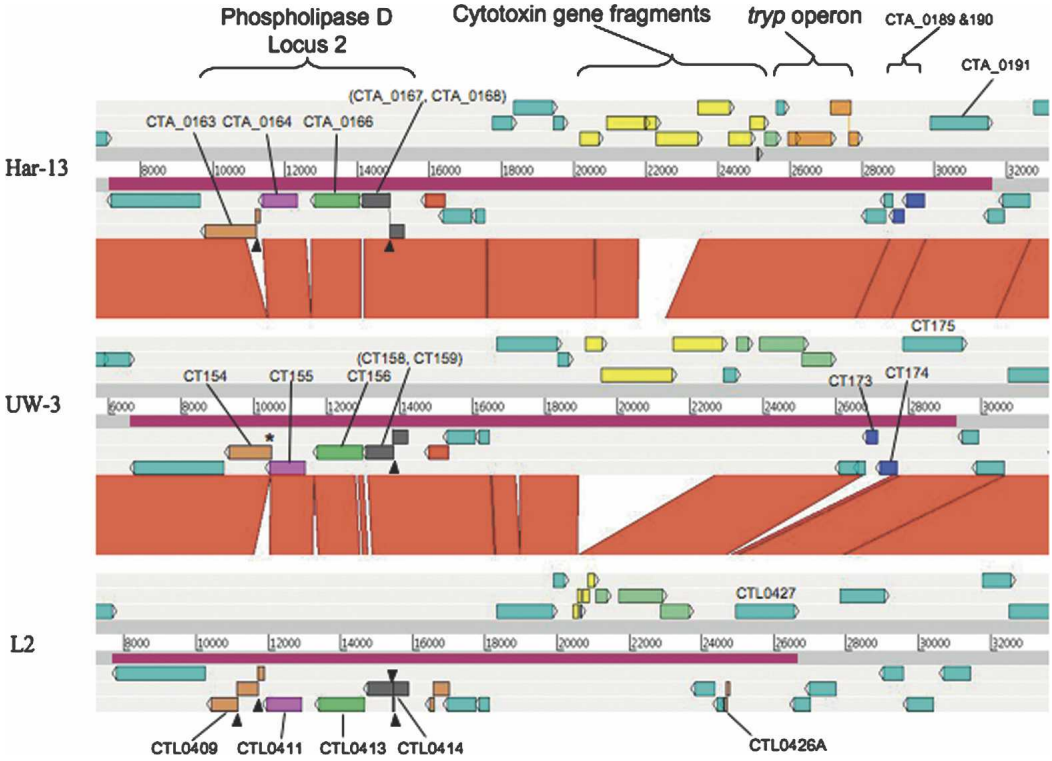

Figure 3. Comparison of the PZ locus of *C. trachomatis* strains L2, UW-3, and Har-13 ACT comparison (http://www.sanger.ac.uk/Software/ACT) of amino acid matches between the complete six-frame translations (computed using TBLASTX) of representatives of the PZ regions of the three sequenced *C. trachomatis* genomes: Har-13, *C. trachomatis* strain Har-13; UW-3, *C. trachomatis* strain UW-3; and L2, *C. trachomatis* strain L2. The red bars spanning between the genomes represent individual TBLASTX matches. Forward and reverse strands of DNA are shown for each genome (dark-gray lines). CDS are marked as colored boxes positioned on the three forward and three reverse translation reading frames (pale-gray lines). Regions mentioned in the text are marked and the CDSs labeled and color coded: Cytotoxin gene fragments (yellow), tryptophan biosynthetic genes (intact pale green; pseudogene orange), and phospholipase D (Locus 2; orthologous PLD CDSs are colored similarly). PZ regions are marked (purple box on DNA lines). The position of deletion events (*) and frameshift mutations (black arrowheads) within the PLD CDS are marked. Multiple systematic gene identifiers contained within parentheses indicates that the CDS was originally annotated as two CDSs.

these two isolates that could explain the differing disease outcomes include altered levels of gene expression brought about, for example, by mutations in the promoter regions or through SNPs that lead to amino acid substitutions that impact on protein function. Whilst it is not possible to accurately predict the former, the possible functional impact of amino acid substitutions can be investigated by comparing the level of conservative and nonconservative sequence changes.

Distribution of SNPs and d_N/d_S analysis of the strain L2 genome

Figure 5 shows that the distribution of SNPs across the genome of strain L2 is not uniform, and that there are discrete regions of high-sequence variation when compared with *C. trachomatis* strains UW-3, Har-13, and UCH-1 (Fig. 5A). The most parsimonious explanation for this is that these regions are hot spots for strong diversifying selection and, therefore, may encode CDSs important for lifestyle. Alternatively, as has been previously suggested for *C. trachomatis,* these candidate loci may act as break points for DNA exchange, thereby introducing the higher levels of sequence variation by recombination (Gomes et al. 2007).

To investigate the functional and evolutionary significance of the identified SNPs, nonsynonymous (d_N) and synonymous (d_S) substitution rates were calculated for strain L2 CDSs in comparison to their orthologs in strains Har-13, UW-3, or UCH-1 (Fig. 5B; Table 3). Comparison of d_N and d_S can be used to assess the influence of selection on protein evolution. A low d_N/d_S ratio (d_N/d_S <<1) indicates strong stabilizing selection, whereas a high ratio (d_N/d_S >1) indicates positive selection and diversification. Since the majority of CDS shared between strain L2 and either Har-13 or UW-3 share >99% sequence identity, it was not possible to accurately identify those CDSs that were truly under strong stabilizing selection. However, the calculation of d_N/d_S was useful to identify genes or domains subject to diversifying selection, such as putative virulence factors or candidate vaccine targets that modulate the host immune response (Smith et al. 1995). Table 3 shows strain L2 CDSs found to have a d_N/d_S ratio >1 in this analysis. The number of CDSs with a d_N/d_S >1 was greatest when strain L2 was compared with for strain UW-3 (63), followed closely by the comparison strain Har-13 (55), and then strain UCH-1 (5).

CDSs, with a d_N/d_S >1, taken from the strains L2 and UCH-1 comparison included *pmpH* (Pmp; polymorphic outer membrane protein), three other membrane proteins, and exodeoxyribonuclease (CTL0583) (data not shown). Similarly, hypothetical and membrane proteins were highly represented in CDSs with a d_N/d_S >1 or both strain Har-13 and UW-3 comparisons (38%–60% depending on the comparison; Table 3). But, also included within proteins showing the highest d_N/d_S >1 were pseudogenes, which would be expected to show a higher level of divergence

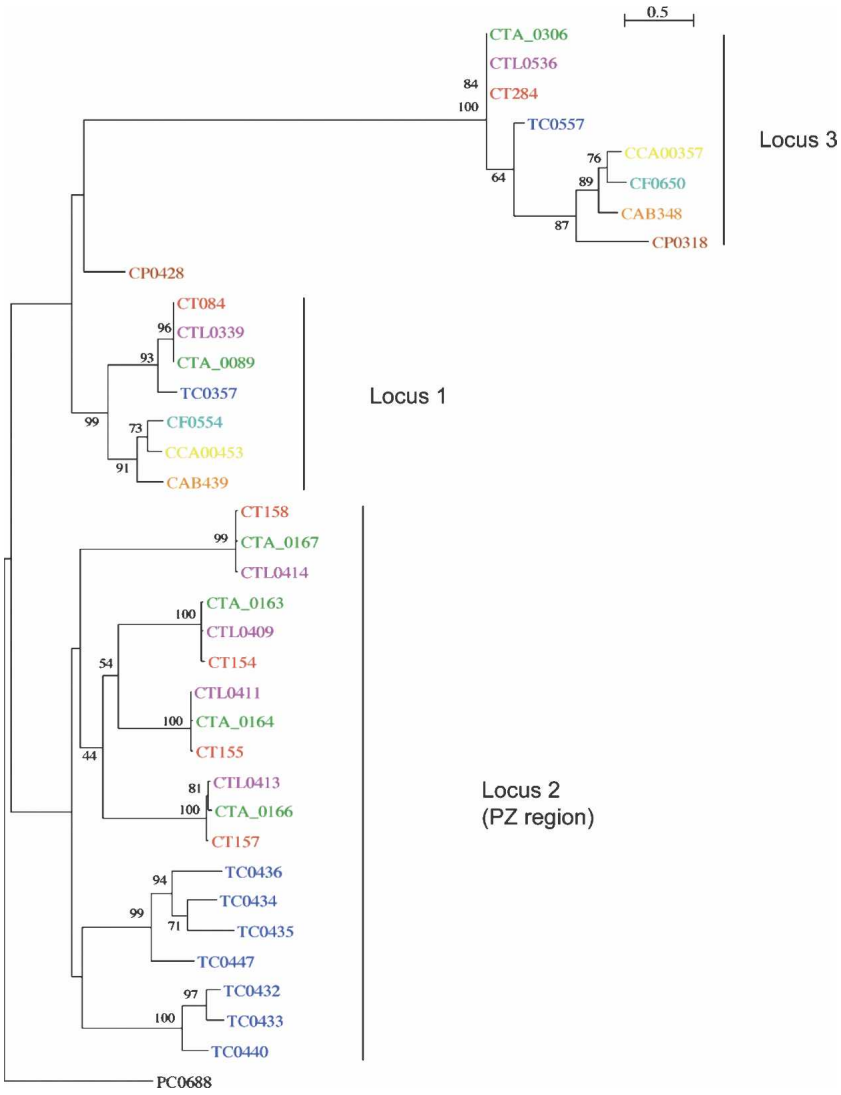

Figure 4. Phylogenetic relationships of chlamydial phospholipase D proteins. The protein names have been colored to indicate chlamydial strains: *C. trachomatis* strains: L2 (pink), UW-3 (red), and Har-13 (green); *C. muridarum* (dark blue); *Cp. felis* (light blue); *Cp. caviae* (yellow); *Cp. abortus* (orange); *Cp. pneumoniae* (brown); *Candidatus Protochlamydia amoebophila* UWE25 (black). Maximum likelihood tree built from protein sequences, using ClustalX, Phylip (Version 3.6), and NJplot . The numbers at the tree branches are percentage bootstrap values indicating the confident levels at that node where congruent. The bar indicates the genetic distance between species (one substitution per 100) as displayed in the branch lengths.

(these genes are not expressed and therefore not subject to purifying selection), and those known to be important for interactions with the host: Pmp proteins (Grimwood and Stephens 1999), inclusion membrane proteins (Inc) (Bannantine et al. 1998a,b, 2000; Rockey et al. 2002) and Tarp. Interestingly, some of the other membrane proteins with a d_N/d_S >1 were predicted to have features of their secondary structure in common with Inc proteins. These have all been denoted candidate Inc (cInc) proteins (Table 3).

The interpretation of d_N/d_S data has to be tempered by the observation of Rocha et al. (2006), who identified a higher preponderance of nonsynonymous substitution in the genomes of closely related strains, as opposed to more distantly related strains. This discrepancy in the rate of substitution is thought to result from the short period of time over which purifying selection has been able to act on the genomes of closely related strains since diversification from a common ancestor. Where nonsynonymous mutations result in amino acid changes that have a mild effect on the fitness of the organism, purifying selection takes longer to exert its evolutionary influence. The observed high d_N/d_S ratio for CDS encoding core metabolic and housekeeping functions (Table 3), which are not obvious targets for diversifying selections, may be due to genetic "noise" rather than a strong selective pressure. To this end, it is worth noting that for some of the CDSs, such as CTL0348 (encoding riboflavin kinase; Table 3), in one comparison the d_N/d_S >1 and in another d_N/d_S <1 (d_N/d_S ratio = 0.725). However, the results shows that many of the CDSs that have high d_N/d_S ratios for both comparisons are from those known to interact with the host, suggesting that this analysis has identified an expanded set of candidate proteins that modulate host–cell interactions.

Discussion

Whole genome sequence analysis of the LGV biovariants completes the data set that now samples the complete range of diversity found in *C. trachomatis*. This work shows that the LGV genome is remarkably similar to the previously sequenced ocular and genital *C. trachomatis* isolates. The low variation in gene content between these biovariants is in agreement with previous microarray data, and rules out the possibility that additional acquired DNA present in *C. trachomatis* strain L2 could explain differences in tissue tropism and disease outcome. It is clear from the comparisons that gene loss and/or small-scale mutational change are the major driving forces shaping host adaptation and tissue tropism of *C. trachomatis*. The major region of variation is the PZ and variation here is primarily due to the loss/degeneration of the cytotoxin gene(s) as previously described (Belland et al. 2001).

In addition, PLDs are important sites of sequence variation, showing both multiple deletions and frameshift mutations. Phospholipases are known to play important roles in pathogenesis in a wide range of bacterial pathogens (Schmiel and Miller 1999; McDermott et al. 2004; McKean et al. 2007). *C. trachomatis* carries three PLD gene clusters, two of which are present in all of the sequenced *Chlamydia* and *Chlamydophila*, and the third cluster, found in the PZ, is specific to *C. trachomatis* and *C. muridarum* and is phylogenetically distinct from the two other PLD clusters. This analysis showed evidence of the PZ PLDs having expanded

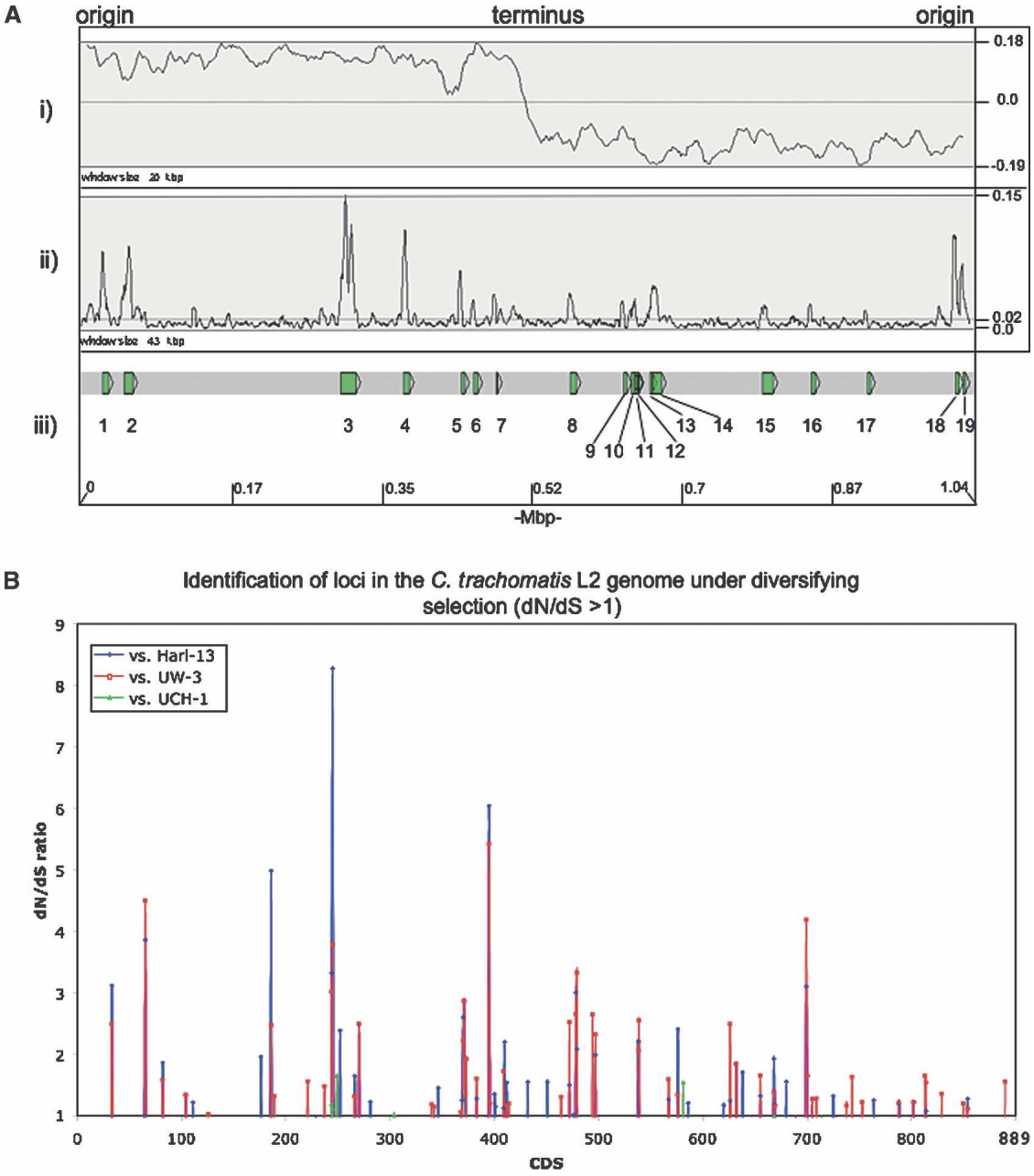

Figure 5. (*A*) Distribution of SNPs over the *C. trachomatis* strain L2 genome compared with strains Har-13, UW-3, and UCH-1. (*i*) The G-C skew (G+C/G-C). The position of the origin and terminus are marked. (*ii*) Shows the sum of all the SNPs in *C. trachomatis* strain L2 compared with strains Har-13, UW-3, and UCH-1, plotted as a frequency over a given window size. (*iii*) Regions with a high SNP density are numbered and the CDSs within this region are described in Supplemental Table 3. The base-pair positions are given in Mbp (*bottom*) and the window sizes are marked. The maximum, minimum, and average frequencies for each plot are given, where appropriate (*right*). (*B*) The distribution of *C. trachomatis* strain L2 CDSs with a d_N/d_S ratio >1. The d_N/d_S values for all of the *C. trachomatis* strain L2 CDSs (*x* axis) compared with their orthologs in strains Har-13, UW-3, and UCH-1 are shown (d_N/d_S >1 only). CDS displaying a high d_N/d_S ratio are described in Table 3.

separately in *C. trachomatis* and *C. muridarum*, suggesting that this family of enzymes may have played, and continue to play, an important role in species-specific adaptation.

These genome comparisons have highlighted evidence for further loss of metabolic capacity. *C. trachomatis* has lost the genes encoding citrate synthase, aconitase, and isocitrate dehydrogenase (McClarty 1999); thus, additional blocks at the level of conversion of succinate to fumarate (*sdhC* pseudogene) and *fumC* in *C. trachomatis* strain L2 would indicate that the TCA cycle does not function in LGV isolates. Interestingly, *Chlamydia* appear to have recently acquired a dicarboxylate transporter (*sodTi*); the role of this gene product seems to be to transport 2-oxoglutarate from the cytoplasm of the host eukaryotic cell. If both fumarase and succinate dehydrogenase are inactive, then 2-oxoglutarate cannot be converted to oxaloacetate, but must be used directly for succinate and succinyl CoA synthesis in the process generating GTP.

Table 3. *C. trachomatis* strain L2 CDSs with a d_N/d_S ratio >1 when compared to either *C. trachomatis* strains Har-13 or UW-3

CDS	d_N/d_S ratio versus Har-13	d_N/d_S ratio versus UW-3	Product description
CTL0247	8.279	3.786	Putative membrane protein
CTL0397	6.049	5.428	Conserved hypothetical protein
CTL0184	4.993	2.477	Candidate inclusion membrane protein (cInc)
CTL0063	3.868	4.499	Conserved hypothetical protein
CTL0246	3.326	3.021	Putative membrane protein
CTL0031	3.125	2.497	Glutamyl-tRNA reductase
CTL0701	3.108	4.192	Cysteine-rich membrane protein
CTL0480	3.01	2.663	Candidate inclusion membrane protein (cInc)
CTL0373	2.866	2.866	Inclusion membrane protein G (IncG)
CTL0372	2.604	2.227	Inclusion membrane protein F (IncF)
CTL0273	2.489	2.494	Putative membrane protein
CTL0578	2.414	1.342	Conserved hypothetical protein (pseudogene)
CTL0255	2.395	2.876	Conserved hypothetical protein
CTL0540	2.216	2.064	Candidate inclusion membrane protein (cInc)
CTL0414	2.207	0.785	Phospholipase D proteins (PLD) (pseudogene)
CTL0481	2.093	3.332	Candidate inclusion membrane protein (cInc)
CTL0499	1.992	2.325	Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex
CTL0173	1.967	0.816	4-Diphosphocytidyl-2-C-methyl-D-erythritol kinase
CTL0670	1.931	1.39	Polymorphic outer membrane protein (PmpB)
CTL0080	1.869	1.58	Conserved hypothetical protein
CTL0633	1.845	1.845	Glucose-6-phosphate isomerase
CTL0269	1.651	1.321	Cytochrome D ubiquinol oxidase subunit II
CTL0702	1.65	1.65	60-kDa cysteine-rich outer membrane protein
CTL0716	1.586	0.948	Translocated actin-recruiting phosphoprotein (Tarp)
CTL0435	1.552	0.619	CTP synthase
CTL0454	1.55	0.461	Oligopeptide transport ATP-binding protein
CTL0417	1.546	1.161	Conserved hypothetical protein
CTL0475	1.505	2.526	Candidate inclusion membrane protein (cInc)
CTL0348	1.459	0.725	Riboflavin kinase
CTL0402	1.354	0.855	Putative integral membrane protein
CTL0102	1.34	1.34	Putative exported protein
CTL0727	1.328	0.993	Two component regulator histidine kinase
CTL0657	1.324	1.655	Lipoamide acyltransferase component (E2)
CTL0374	1.317	1.317	Inclusion membrane protein A
CTL0385	1.287	1.608	ABC transporter, ATP-binding component
CTL0859	1.277	1.116	Thiol:disulfide interchange protein
CTL0707	1.272	1.272	Single-stranded DNA-specific exonuclease
CTL0569	1.269	1.596	LSU ribosomal protein L10P
CTL0767	1.259	0.628	Glyceraldehydes 3-phosphate dehydrogenase
CTL0626	1.245	2.494	Conserved hypothetical protein
CTL0284	1.229	0.306	Ribonuclease HII
CTL0805	1.223	1.223	Histidyl-tRNA synthetase
CTL0109	1.221	0.609	Na^+-translocating NADH-quinone reductase subunit F
CTL0587	1.209	0.592	Excinuclease ABC subunit A
CTL0398	1.208	1.208	Conserved hypothetical protein
CTL0791	1.198	1.214	Putative membrane protein
CTL0671	1.185	1.152	Polymorphic outer membrane protein (PmpC)
CTL0620	1.17	0.667	3-Phosphoshikimate 1-carboxyvinyltransferase
CTL0403	1.148	0.234	FAD-dependent monooxygenase
CTL0413	1.125	1.724	Phospholipase D proteins (PLD)
CTL0818	1.08	1.08	Putative helicase (SWF/SNF family)
CTL0478	1.02	0.461	Candidate inclusion membrane protein (cInc)
CTL0062	1.006	0.669	Phosphoglycerate kinase
CTL0344	0.983	1.146	Low-calcium response protein E (TTSS effector protein)
CTL0746	0.978	1.635	Ferrochelatase
CTL0219	0.931	1.56	Conserved hypothetical protein
CTL0496	0.909	2.652	Conserved hypothetical protein
CTL0854	0.903	1.2	Succinate dehydrogenase iron-sulfur protein
CTL0370	0.856	1.065	Inclusion membrane protein D (IncD)
CTL0187	0.823	1.318	Phosphoglucosamine mutase
CTL0756	0.733	1.225	Putative nucleotide transport protein
CTL0124	0.684	1.028	60-kDa chaperonin GroEL
CTL0833	0.678	1.357	General secretion pathway protein F
CTL0235	0.523	1.477	Conserved hypothetical protein
CTL0894	0.466	1.558	Transcriptional regulatory protein
CTL0342	0.296	1.187	4-Alpha-glucanotransferase
CTL0741	0.292	1.168	Oligopeptide-transport protein, binding subunit

The significance of finding mutations in *arcD* and *PvlArgDC* in strains L2 and Har-13, respectively, whilst they are apparently intact in the genital isolate strain UW-3, is unclear. Studies on the pH of *C. trachomatis* inclusions indicates that either a similar homeostatic balance to the host cell (Dautry-Varsat et al. 2005) or a pH close to 6.0 occurs; thus, it seems unlikely that the presence of intact arginine transport and metabolism alone accounts for the differences in disease outcomes associated with the different genital tract biovariants (Schramm et al. 1996).

Whilst it is clear from experimental data that loss of a particular pathway (e.g., *trp* biosynthesis) accounts for a single, important, metabolically defined difference between chlamydial biovariants (Stephens et al. 1998; Fehlner-Gardiner et al. 2002; Carlson et al. 2005), this is likely to be a unique observation, as there are no other single genetic features that correlate with a particular biotype and for which there is also phenotypic evidence. Thus, the new genome sequence data shows that the determinants of tropism and invasiveness of LGV isolates are likely to be multifactorial and complex. Without a genetic system to allow testing of the properties of "single gene" mutations, it is not possible to assign pathogenic properties to single factors.

The most overwhelming impression of the *C. trachomatis* strain L2 genome is the lack of variation in terms of coding capacity when compared with the other serovars. This is further reinforced by the comparison of strain L2 with the recent proctitis strain UCH-1 (isolated in 2006). It has been proposed that the LGV proctitis strain is a new epidemic isolate that is rapidly spreading worldwide. Whilst the symptoms caused by the proctitis strains are atypical, the genome of UCH-1 is almost identical to that of strain L2, which was isolated in 1969 and has been adopted as a "model" strain by many laboratories for routine studies. Although it is impossible to rule out that SNPs do not subtly change the function or the level of expression of key genes within the genome, we can say that there is no additional coding capacity to explain the differences in clinical manifestations. We favor the views of Schachter and Monocada (2005) that this strain, far from being a newly introduced and rapidly spreading infectious strain, is simply a classical LGV isolate and has been circulating in the human population for a long time: an old strain causing a new disease.

Methods

Bacterial strains

We sequenced the widely studied *C. trachomatis* strain L2/434/Bu (L2; ATCC VR902B) and a recent clinical isolate strain L2/UCH-1/proctitis (UCH-1). *C. trachomatis* strain UCH-1 was isolated from a rectal swab of a 49-yr-old MSM who was HIV positive and Hepatitis C negative. He attended a London Genital and Urinary Medicine (GUM) clinic in January 2006 showing clinical signs of proctitis. The rectal sample tested positive for *C. trachomatis* (negative for *Neisseria gonorrhoeae*) using a conventional *C. trachomatis* Nucleic Acid Amplifcation Test (CT NAAT). This sample was sent to the Health Protection Agency (Colindale) for routine LGV analysis. Confirmation that this strain was a LGV biovariant was ascertained using the primers CT1 and CT5 as described in Ngandjio et al. (2003). To determine whether this isolate was part of the wider European outbreak of the newly described L2b group of LGV isolates, we sequenced the *ompA* gene and identified the definitive base change at nucleotide 485. No other variations in the *ompA* gene were detected, indicating that the sample contained a single isolate and, hence, it was not considered necessary to perform plaque purification. This minimized the amount of passaging in vitro and the likelihood that attenuating mutations would arise during culture.

Growth of *C. trachomatis* strains L2 and UCH-1

C. trachomatis strain L2 was propagated in L929 cells in suspension cultures and the DNA purified from EBs as previously described (Birkelund and Stephens 1992).

C. trachomatis strain UCH-1 was initially isolated in BGMK cells in Dulbecco's Modified Eagle's Medium (DMEM) containing 0.03 M glucose, vancomycin at 10 $\mu g/mL^{-1}$, Gentamicin at 10 $\mu g/mL^{-1}$, and Cycloheximide at 1 $\mu g/mL^{-1}$.

Purification of chromosomal DNA for genome sequencing

Mycoplasma-free BGMK cells were used to amplify *C. trachomatis* strain UCH-1. Six passages in vitro were sufficient to yield 14 × T-175 tissue culture flasks infected to confluence with this strain. EBs, purified as described by Skipp et al. (2005), were incubated with protease K (200 µg/mL) for 1 h at 60°C prior to extraction of chromosomal DNA using a Promega Wizard Genomic Purification kit according to the manufacturer's protocol. The *C. trachomatis* strain UCH-1 genomic DNA was quantified spectrophotometrically and analyzed by gel electrophoresis prior to genome sequence analysis.

Sequencing *C. trachomatis* strains L2 and UCH-1

The genome of strain L2 was obtained from 21,573 end sequences (giving 8× coverage) derived from pUC18 (insert size 1.4–2.2 kb) small insert libraries using dye terminator chemistry on ABI3700 automated sequencers. End sequences from larger insert plasmid (pMAQ1 9–12 kb, 9–12-kb insert size) libraries were used as a scaffold. The sequence was assembled, finished, and annotated as described previously (Parkhill et al. 2000) using the program Artemis (Berriman and Rutherford 2003) to collate data and facilitate annotation.

C. trachomatis strain UCH-1 chromosomal DNA was sequenced using a 454/Roche GS20 machine, according to the manufacturer's protocols, producing 304,953 reads with an average length of 105 bp, representing a theoretical 30.8-fold coverage of the genome. The 454 sequence reads were assembled de novo into 18 nonredundant contigs with an average of 29.8-fold coverage, using the 454/Roche Newbler assembly program. These contigs were reordered based on BLAST alignments with strain L2. The gaps between these contigs were closed by directed PCR and the products sequenced with BigDye terminator chemistry on ABI3730 capillary sequencers.

The genomes of *C. trachomatis* strains L2 and UCH-1 have been submitted to the public database.

In silico genome analysis

The genome sequences of *C. trachomatis* strains L2, UW-3, and Har-13 were compared pairwise using the Artemis Comparison Tool (ACT) (Carver et al. 2005). Pseudogenes had one or more mutations that would ablate expression; each of the inactivating mutations was subsequently checked against the original sequencing data.

Orthologous gene sets were identified by reciprocal FASTA searches. Only those pairs of homologous CDSs were retained for further analysis, where the predicted amino acid identity was ≥40% over 80% of the protein length. These genes were then subject to manual curation using gene synteny to increase the accuracy of this analysis. This strategy was applied to pairwise comparisons of the genomes of *C. trachomatis* strains L2, UCH-1, UW-3, and Har-13, as well as *C. muridarum* (strain Nigg), *Cp. felis*

(strain Fe/C-56), *Cp. caviae* (strain GPIC), and *Cp. pneumoniae* (strain AR39).

Synonymous and nonsynonymous substitution-rate calculations were determined for orthologous gene sets generated by reciprocal best-match FASTA analysis and aligned using Needle from the EMBOSS suit of software (Rice et al. 2000). d_N and d_S values were calculated using the PAML software suit (Yang 1997) by the method of Nei and Gojobori (1986), as implemented by the yn00 program (part of the PAML software suit). The method developed by Nei and Gojobori uses the Jukes and Cantor (1969) formula to estimate the number of synonymous and nonsynonymous substitutions per site.

Acknowledgments

We thank the core-sequencing and informatics teams at the Sanger Institute for their assistance and The Wellcome Trust for its support of the Sanger Institute Pathogen Sequencing Unit. This work was supported by The Wellcome Trust grants 080348 to I.N.C. We acknowledge the expert technical support of Lesley Cutcliffe, Rachel Skilton, S. Kalman, and R.W. Davis.

References

Azuma, Y., Hirakawa, H., Yamashita, A., Cai, Y., Rahman, M.A., Suzuki, H., Mitaku, S., Toh, H., Goto, S., Murakami, T., et al. 2006. Genome sequence of the cat pathogen, *Chlamydophila felis*. *DNA Res*. **13:** 15–23.

Bannantine, J.P., Rockey, D.D., and Hackstadt, T. 1998a. Tandem genes of *Chlamydia psittaci* that encode proteins localized to the inclusion membrane. *Mol. Microbiol*. **28:** 1017–1026.

Bannantine, J.P., Stamm, W.E., Suchland, R.J., and Rockey, D.D. 1998b. *Chlamydia trachomatis* IncA is localized to the inclusion membrane and is recognized by antisera from infected humans and primates. *Infect. Immun*. **66:** 6017–6021.

Bannantine, J.P., Griffiths, R.S., Viratyosin, W., Brown, W.J., and Rockey, D.D. 2000. A secondary structure motif predictive of protein localization to the chlamydial inclusion membrane. *Cell. Microbiol*. **2:** 35–47.

Behets, F.M., Andriamiadana, J., Randrianasolo, D., Randriamanga, R., Rasamilalao, D., Chen, C.Y., Weiss, J.B., Morse, S.A., Dallabetta, G., and Cohen, M.S. 1999. Chancroid, primary syphilis, genital herpes, and lymphogranuloma venereum in Antananarivo, Madagascar. *J. Infect. Dis*. **180:** 1382–1385.

Belland, R.J., Scidmore, M.A., Crane, D.D., Hogan, D.M., Whitmire, W., McClarty, G., and Caldwell, H.D. 2001. *Chlamydia trachomatis* cytotoxicity associated with complete and partial cytotoxin genes. *Proc. Natl. Acad. Sci*. **98:** 13984–13989.

Berriman, M. and Rutherford, K. 2003. Viewing and annotating sequence data with Artemis. *Brief. Bioinform*. **4:** 124–132.

Birkelund, S. and Stephens, R.S. 1992. Construction of physical and genetic maps of *Chlamydia trachomatis* serovar L2 by pulsed-field gel electrophoresis. *J. Bacteriol*. **174:** 2742–2747.

Caldwell, H.D., Wood, H., Crane, D., Bailey, R., Jones, R.B., Mabey, D., Maclean, I., Mohammed, Z., Peeling, R., Roshick, C., et al. 2003. Polymorphisms in *Chlamydia trachomatis* tryptophan synthase genes differentiate between genital and ocular isolates. *J. Clin. Invest*. **111:** 1757–1769.

Carlson, J.H., Hughes, S., Hogan, D., Cieplak, G., Sturdevant, D.E., McClarty, G., Caldwell, H.D., and Belland, R.J. 2004. Polymorphisms in the *Chlamydia trachomatis* cytotoxin locus associated with ocular and genital isolates. *Infect. Immun*. **72:** 7063–7072.

Carlson, J.H., Porcella, S.F., McClarty, G., and Caldwell, H.D. 2005. Comparative genomic analysis of *Chlamydia trachomatis* oculotropic and genitotropic strains. *Infect. Immun*. **73:** 6407–6418.

Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.A., Barrell, B.G., and Parkhill, J. 2005. ACT: The Artemis comparison tool. *Bioinformatics* **21:** 3422–3423.

Comanducci, M., Ricci, S., and Ratti, G. 1988. The structure of a plasmid of *Chlamydia trachomatis* believed to be required for growth within mammalian cells. *Mol. Microbiol*. **2:** 531–538.

Dautry-Varsat, A., Subtil, A., and Hackstadt, T. 2005. Recent insights into the mechanisms of *Chlamydia* entry. *Cell. Microbiol*. **7:** 1714–1722.

Delcher, A.L., Kasif, S., Fleischmann, R.D., Peterson, J., White, O., and Salzberg, S.L. 1999. Alignment of whole genomes. *Nucleic Acids Res*. **27:** 2369–2376. doi: 10.1093/nar/27.11.2369.

Fehlner-Gardiner, C., Roshick, C., Carlson, J.H., Hughes, S., Belland, R.J., Caldwell, H.D., and McClarty, G. 2002. Molecular basis defining human *Chlamydia trachomatis* tissue tropism. A possible role for tryptophan synthase. *J. Biol. Chem*. **277:** 26893–26903.

Gomes, J.P., Bruno, W.J., Nunes, A., Santos, N., Florindo, C., Borrego, M.J., and Dean, D. 2007. Evolution of *Chlamydia trachomatis* diversity occurs by widespread interstrain recombination involving hotspots. *Genome Res*. **17:** 50–60.

Graham, D.E., Xu, H., and White, R.H. 2002. *Methanococcus jannaschii* uses a pyruvoyl-dependent arginine decarboxylase in polyamine biosynthesis. *J. Biol. Chem*. **277:** 23500–23507.

Grimwood, J. and Stephens, R.S. 1999. Computational analysis of the polymorphic membrane protein superfamily of *Chlamydia trachomatis* and *Chlamydia pneumoniae*. *Microb. Comp. Genomics* **4:** 187–201.

Guest, J.R. and Roberts, R.E. 1983. Cloning, mapping, and expression of the fumarase gene of *Escherichia coli* K-12. *J. Bacteriol*. **153:** 588–596.

Guest, J.R., Miles, J.S., Roberts, R.E., and Woods, S.A. 1985. The fumarase genes of *Escherichia coli:* Location of the *fumB* gene and discovery of a new gene (f*umC*). *J. Gen. Microbiol*. **131:** 2971–2984.

Jewett, T.J., Fischer, E.R., Mead, D.J., and Hackstadt, T. 2006. Chlamydial TARP is a bacterial nucleator of actin. *Proc. Natl. Acad. Sci*. **103:** 15599–15604.

Jukes, T.H. and Cantor, R.C. 1969. Evolution of protein molecules. In *Mammalian protein metabolism III.* (ed. H.N. Munro) pp. 21–132. Academic Press, New York.

Kieboom, J. and Abee, T. 2006. Arginine-dependent acid resistance in *Salmonella enterica* serovar Typhimurium. *J. Bacteriol*. **188:** 5650–5653.

Longbottom, D. and Coulter, L.J. 2003. Animal chlamydioses and zoonotic implications. *J. Comp. Pathol*. **128:** 217–244.

Mabey, D. and Peeling, R.W. 2002. Lymphogranuloma venereum. *Sex. Transm. Infect*. **78:** 90–92.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437:** 376–380.

McClarty, G. 1999. Chlamydial metabolism as inferred from the complete genome sequence. In *Chlamydia: Intracellular biology,pathogenesis, and immunity*. pp. 69–100. Wiley, New York.

McDermott, M., Wakelam, M.J., and Morris, A.J. 2004. Phospholipase D. *Biochem. Cell Biol*. **82:** 225–253.

McKean, S.C., Davies, J.K., and Moore, R.J. 2007. Expression of phospholipase D, the major virulence factor of *Corynebacterium pseudotuberculosis*, is regulated by multiple environmental factors and plays a role in macrophage death. *Microbiol*. **153:** 2203–2211.

Nei, M. and Gojobori, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol*. **3:** 418–426.

Nelson, D.E., Crane, D.D., Taylor, L.D., Dorward, D.W., Goheen, M.M., and Caldwell, H.D. 2006. Inhibition of chlamydiae by primary alcohols correlates with the strain-specific complement of plasticity zone phospholipase D genes. *Infect. Immun*. **74:** 73–80.

Ngandjio, A., Clerc, M., Fonkoua, M.C., Thonnon, J., Njock, F., Pouillot, R., Lunel, F., Bebear, C., De Barbeyrac, B., and Bianchi, A. 2003. Screening of volunteer students in Yaounde (Cameroon, Central Africa) for *Chlamydia trachomatis i*nfection and genotyping of isolated *C. trachomatis* strains. *J. Clin. Microbiol*. **41:** 4404–4407.

Nieuwenhuis, R.F., Ossewaarde, J.M., van der Meijden, W.I., and Neumann, H.A. 2003. Unusual presentation of early lymphogranuloma venereum in an HIV-1 infected patient: Effective treatment with 1 g azithromycin. *Sex. Transm. Infect*. **79:** 453–455.

Parkhill, J., Wren, B.W., Mungall, K., Ketley, J.M., Churcher, C., Basham, D., Chillingworth, T., Davies, R.M., Feltwell, T., Holroyd, S., et al. 2000. The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* **403:** 665–668.

Peeling, R.W. and Brunham, R.C. 1996. Chlamydiae as pathogens: New species and new issues. *Emerg. Infect. Dis*. **2:** 307–319.

Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Utterback, T., Berry, K., et al. 2000. Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Res*. **28:** 1397–1406. doi: 10.1093/nar/28.6.1397.

Read, T.D., Myers, G.S., Brunham, R.C., Nelson, W.C., Paulsen, I.T., Heidelberg, J., Holtzapple, E., Khouri, H., Federova, N.B., Carty, H.A., et al. 2003. Genome sequence of *Chlamydophila caviae* (*Chlamydia psittaci* GPIC): Examining the role of niche-specific genes in the

evolution of the *Chlamydiaceae*. *Nucleic Acids Res.* **31:** 2134–2147. doi: 10.1093/nar/gkg321.
Rice, P., Longden, I., and Bleasby, A. 2000. EMBOSS: The european molecular biology open software suite. *Trends Genet.* **16:** 276–277.
Rocha, E.P., Smith, J.M., Hurst, L.D., Holden, M.T., Cooper, J.E., Smith, N.H., and Feil, E.J. 2006. Comparisons of d_N/d_S are time dependent for closely related bacterial genomes. *J. Theor. Biol.* **239:** 226–235.
Rockey, D.D., Scidmore, M.A., Bannantine, J.P., and Brown, W.J. 2002. Proteins in the chlamydial inclusion membrane. *Microbes Infect.* **4:** 333–340.
Saikku, P. 1999. Epidemiology of *Chlamydia pneumoniae* in atherosclerosis. *Am. Heart J.* **138:** S500–S503. doi: 10.1016/S0002-8703(99)70285-1.
Schachter, J. and Moncada, J. 2005. Lymphogranuloma venereum: How to turn an endemic disease into an outbreak of a new disease? Start looking. *Sex. Transm. Dis.* **32:** 331–332.
Schmiel, D.H. and Miller, V.L. 1999. Bacterial phospholipases and pathogenesis. *Microbes Infect.* **1:** 1103–1112.
Schramm, N., Bagnell, C.R., and Wyrick, P.B. 1996. Vesicles containing *Chlamydia trachomatis* serovar L2 remain above pH 6 within HEC-1B cells. *Infect. Immun.* **64:** 1208–1214.
Shaw, A.C., Gevaert, K., Demol, H., Hoorelbeke, B., Vandekerckhove, J., Larsen, M.R., Roepstorff, P., Holm, A., Christiansen, G., and Birkelund, S. 2002. Comparative proteome analysis of *Chlamydia trachomatis* serovar A, D and L2. *Proteomics* **2:** 164–186.
Skipp, P., Robinson, J., O'Connor, C.D., and Clarke, I.N. 2005. Shotgun proteomic analysis of *Chlamydia trachomatis*. *Proteomics* **5:** 1558–1573.
Smith, N.H., Maynard Smith, J., and Spratt, B.G. 1995. Sequence evolution of the *porB* gene of *Neisseria gonorrhoeae* and *Neisseria meningitidis*: Evidence of positive Darwinian selection. *Mol. Biol. Evol.* **12:** 363–370.
Spaargaren, J., Fennema, H.S., Morre, S.A., de Vries, H.J., and Coutinho, R.A. 2005. New lymphogranuloma venereum *Chlamydia trachomatis* variant, Amsterdam. *Emerg. Infect. Dis.* **11:** 1090–1092.
Stephens, R.S., Kalman, S., Lammel, C., Fan, J., Marathe, R., Aravind, L., Mitchell, W., Olinger, L., Tatusov, R.L., Zhao, Q., et al. 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science* **282:** 754–759.
Thomson, N.R., Yeats, C., Bell, K., Holden, M.T., Bentley, S.D., Livingstone, M., Cerdeno-Tarraga, A.M., Harris, B., Doggett, J., Ormond, D., et al. 2005. The *Chlamydophila abortus* genome sequence reveals an array of variable proteins that contribute to interspecies variation. *Genome Res.* **15:** 629–640.
Viravan, C., Dance, D.A., Ariyarit, C., Looareesuwan, S., Wattanagoon, Y., Davis, T.M., Wuthiekanun, V., Tantivanich, S., Angus, B.J., and White, N.J. 1996. A prospective clinical and bacteriologic study of inguinal buboes in Thai men. *Clin. Infect. Dis.* **22:** 233–239.
Yang, Z. 1997. PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13:** 555–556.

Received August 10, 2007; accepted in revised form October 15, 2007.