Infection, Genetics and Evolution 32 (2015) 74–88

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Deep comparative genomics among *Chlamydia trachomatis* lymphogranuloma venereum isolates highlights genes potentially involved in pathoadaptation

Vítor Borges, João Paulo Gomes *

Reference Laboratory of Bacterial Sexually Transmitted Infections, Department of Infectious Diseases, National Institute of Health, Av. Padre Cruz, 1649-016 Lisbon, Portugal
Bioinformatics Unit, Department of Infectious Diseases, National Institute of Health, Av. Padre Cruz, 1649-016 Lisbon, Portugal

ARTICLE INFO

Article history:
Received 17 December 2014
Received in revised form 23 February 2015
Accepted 26 February 2015
Available online 3 March 2015

Keywords:
Chlamydia trachomatis
Lymphogranuloma venereum
Proctitis
Genetic variability
Pathoadaptation

ABSTRACT

Lymphogranuloma venereum (LGV) is a human sexually transmitted disease caused by the obligate intracellular bacterium *Chlamydia trachomatis* (serovars L1–L3). LGV clinical manifestations range from severe ulcerative proctitis (anorectal syndrome), primarily caused by the epidemic L2b strains, to painful inguinal lymphadenopathy (the typical LGV bubonic form). Besides potential host-related factors, the differential disease severity and tissue tropism among LGV strains is likely a function of the genetic backbone of the strains. We aimed to characterize the genetic variability among LGV strains as strain- or serovar-specific mutations may underlie phenotypic signatures, and to investigate the mutational events that occurred throughout the pathoadaptation of the epidemic L2b lineage. By analyzing 20 previously published genomes from L1, L2, L2b and L3 strains and two new genomes from L2b strains, we detected 1497 variant sites and about 100 indels, affecting 453 genes and 144 intergenic regions, with 34 genes displaying a clear overrepresentation of nonsynonymous mutations. Effectors and/or type III secretion substrates (almost all of those described in the literature) and inclusion membrane proteins showed amino acid changes that were about fivefold more frequent than silent changes. More than 120 variant sites occurred in plasmid-regulated virulence genes, and 66% yielded amino acid changes. The identified serovar-specific variant sites revealed that the L2b-specific mutations are likely associated with higher fitness and pointed out potential targets for future highly discriminatory diagnostic/typing tests. By evaluating the evolutionary pathway beyond the L2b clonal radiation, we observed that 90.2% of the intra-L2b variant sites occurring in coding regions involve nonsynonymous mutations, where CT456/*tarp* has been the main target. Considering the progress on *C. trachomatis* genetic manipulation, this study may constitute an important contribution for prioritizing study targets for functional genomics aiming to dissect the impact of the identified intra-LGV polymorphisms on virulence or tropism dissimilarities among LGV strains.

1. Introduction

Lymphogranuloma venereum (LGV) is a human sexually transmitted disease caused by strains from the serovars L1–L3 of the obligate intracellular bacterium *Chlamydia trachomatis*. Upon genital or rectal contact, these strains have the ability to infect mononuclear phagocytes and spread into regional lymph nodes causing an inguinal syndrome characterized by genital ulcers and painful inguinal lymphadenopathy (the typical LGV bubonic form) (Mabey and Peeling, 2002; Thomson et al., 2008). Therefore, LGV strains have been historically considered to represent a biovar (LGV biovar) distinct from the trachoma biovar that is comprised by strains from the ocular serovars (A–C) and genital serovars (D–K), which otherwise are normally associated with infections restricted to the epithelial cells of the conjunctival and genital mucosae, respectively. The ocular strains are the causative agent of trachoma, the world's leading cause of preventable infectious blindness (Wright et al., 2008), whereas *C. trachomatis* strains from serovars D–K are the major cause of bacterial sexually transmitted diseases worldwide (Bébéar and de Barbeyrac, 2009).

Although cases of typical bubonic LGV presentation have been endemic in parts of Africa, Asia, South America and the Caribbean but are rare in the Western World (Bauwens et al.,

* Corresponding author at: Department of Infectious Diseases, National Institute of Health, Av. Padre Cruz, 1649-016 Lisbon, Portugal. Tel.: +351 217 519 241; fax: +351 217 526 400.
E-mail address: j.paulo.gomes@insa.min-saude.pt (J.P. Gomes).

http://dx.doi.org/10.1016/j.meegid.2015.02.026
1567-1348/

2002; Behets et al., 1999; Dal Conte et al., 2014; Mabey and Peeling, 2002; O'Farrell et al., 1991; Viravan et al., 1996; White, 2009), an atypical LGV clinical presentation characterized by severe ulcerative proctitis (so called anorectal syndrome) has emerged in Europe and North America since 2003 (Dal Conte et al., 2014; de Vrieze and de Vries, 2014; Nieuwenhuis et al., 2003; van de Laar, 2006; Savage et al., 2009; White, 2009). This anorectal syndrome, which mostly afflicts men who have sex with men (MSM), normally co-infected with HIV and other sexually transmitted infections (de Vries et al., in press; Savage et al., 2009), was found to be primarily caused by strains from the LGV serovar L2b. The genome sequencing of epidemic L2b isolates causing proctitis revealed that these isolates are nearly clonal (Harris et al., 2012), and carry no evident genomic particularities (e.g., gene gain/loss) that could unequivocally be associated with the atypical symptoms (Thomson et al., 2008). Nonetheless, although sporadically, L2b infections may progress to inguinal lymphadenopathy or persistent genital ulcers (de Lavaissière and Nougué, 2013; Marcotte et al., 2014; Savage et al., 2009; White, 2009), and L2b-associated cases of proctitis and bubonic LGV in a woman have been described (de Vries et al., in press; Peuchant et al., 2011; Verweij et al., 2012), suggesting that L2b strains may carry specific genetic features that enable them to exhibit wider tropism and transmission skills than strains from the remainder LGV serovars. These data, together with the existence of LGV strains with rather unusual recombinant profiles (Somboonna et al., 2011), clearly suggest that the diversity among LGV strains is higher than expected (Harris et al., 2012), and thus, point out that more sampling and deeper molecular characterization of LGV strains are needed to better understand the pathodiversity within the LGV biovar. For instance, comparative genomic analysis of *C. trachomatis* ocular strains was important to find subtle genomic variations likely underlying virulence dissimilarities (Kari et al., 2008), and we have previously identified some genes targeted by positive selection events likely driving phenotypic diversity among LGV strains (Borges et al., 2012). Besides host-related factors (genetics and/or immunological) [reviewed in Abdelsamed et al. (2013) and Asner et al. (in press)], the differential disease severity and tissue tropism among LGV strains is likely a function of the genetic backbone of the strain, where few strain- or serovar-specific mutations may be determinant factors for particular phenotypic signatures. In fact, although LGV strains are considerably more related than those of the trachoma biovar and discreet recombination within the LGV biovar has occurred, hundreds of mutations are known to be fixed since the LGV ancestral lineage diverged from the remaining *C. trachomatis* strains (Borges et al., 2012; Harris et al., 2012; Joseph et al., 2012). In addition, the evolutionary pathway underlying the L2b clonal expansion and dispersion in MSM is still to be deciphered. Considering this scenario, we intended to deeply characterize the genetic variability among LGV strains and to get insight into the mutational events that occurred throughout the pathoadaptation of the epidemic L2b lineage.

2. Materials and methods

2.1. C. trachomatis strains, cell culture and bacterial DNA purification

Two serovar L2b *C. trachomatis* clinical strains [*ompA* genotype confirmed as previously described (Nunes et al., 2009)] were isolated from positive diagnostic specimens collected from MSM suspected of having LGV proctitis that attended to the major Portuguese sexually transmitted disease clinic (Lapa Health Centre, Lisbon). The L2b/CS19/08 strain was isolated from a anoarectal swab of a 29-years old MSM with proctitis, syphilis and HIV(+), whereas the L2b/CS784/08 strain was collected from an anoarectal swab of a 32-years old MSM having anal bloody discharge and being both HIV and *Treponema pallidum* negative. Each clinical specimen in transport medium (2 sucrose phosphate buffer supplemented with gentamicin, vancomycin and nistatin) was inoculated (0.2 ml per well) in HeLa229 confluent monolayers (cultured on 24-well plates in MEM containing 10% fetal bovine serum, 5 mM L-glutamic acid, 10 μg/ml gentamicin and 0.5 μg/ml fungizone at 37 °C, 5% CO_2) by centrifuging at 3500 rpm for 1 h at 34 °C. The cultures were subsequently incubated for 1 h at 37 °C, 5% CO_2, and the cell medium was replaced by fresh medium supplemented with vitamins (1x), non-essential aminoacids (1×), glucose (0.5%) and cycloheximide (0.5 μg/ml). Cultures were allowed to grow at 37 °C, 5%CO_2 until about 48 h post-infection. The yield of infection was monitored by immunofluorescence microscopy after fixing cultures with methanol and staining with an anti-*C. trachomatis* lipopolysaccharide antibody (Pathfinder), according to manufacturer's instructions. The medium was then removed, the bacterial-infected cells were harvested using glass beads and disrupted through sonication (Vibra Cell, Bioblock Scientific), and further submitted to low-speed centrifugation (700 rpm for 7 min). The bacterial-enriched supernatant was collected, homogenized and re-inoculated onto new HeLa229 confluent monolayers, as described above. Both chlamydial cultures were scale-up (6–7 passages) until obtaining eight T25 cm2 flasks containing chlamydial-enriched HeLa229 monolayers, and further subjected to a discontinuous density gradient purification procedure, as previously described (Borges et al., 2013). DNA was finally extracted from the elementary bodies (EBs) fraction using the DNA Mini Kit (Qiagen, Valencia, CA, USA) according to manufacturer's instructions, and further quantified by applying the Quant-iT™ PicoGreen® dsDNA Assay (Invitrogen). The DNA integrity was assessed through agarose gel electrophoresis.

2.2. Whole-genome sequencing, assembly and alignment

High-quality DNA samples were used to prepare Illumina paired-end libraries according to the manufacturer's instructions (by using Nextera XT DNA sample preparation and index kits; Illumina Inc., San Diego, CA). DNA libraries were further loaded onto MiSeq reagent cartridge (MiSeq Reagent Kit v2; Illumina Inc., San Diego, CA, USA) and, subsequently, subjected to cluster generation and paired-end sequencing (2 × 250 bp) on a MiSeq platform (Illumina Inc., San Diego, CA, USA), according to the manufacturer's instructions. Illumina reads were mapped to chromosome and plasmid sequences from the *C. trachomatis* strain L2b/UCH-1 (GenBank accession numbers AM884177 and AM886279) (Thomson et al., 2008) using both Bowtie2 (version 2.1.0 [http://bowtie-bio.sourceforge.net/bowtie2/index.shtml]) (Langmead and Salzberg, 2012) and the Burrows-Wheeler Aligner (BWA) software (version 0.7.5a [http://bio-bwa.sourceforge.net/]) (Li and Durbin, 2010). The reference-based approach seemed accurate as the epidemic L2b isolates are expectedly clonal and carry no gene gain/loss differences (Thomson et al., 2008; Harris et al., 2012). So, the chance of finding new, unique sequences for these two new genomes that did not align to L2b/UCH-1 is virtually zero. The obtained mean depth coverage for the L2/CS19/08 and L2b/CS784/08 strains were 68-fold and 26-fold (for the chromosome) and 586-fold and 226-fold (for the plasmid), respectively; with no single regions displaying zero coverage. Based on the ratio plasmid/chromosome taken from the respective depth coverage, we were able to infer the plasmid copy number *per* chromosome (about nine for both strains). SAMtools/BCFtools (http://samtools.sourceforge.net/) (Li et al., 2009) were applied to call Single Nucleotide Polymorphisms (SNPs) and indels, which were confirmed through visual inspection using the Integrative Genomics Viewer (version 2.3.12 [http://www.broadinstitute.org/igv])