Universidade Nova de Lisboa

Instituto de Higiene e Medicina Tropical

Propagation and Molecular Characterization of *Chlamydia trachomatis* strains isolated in Portugal

Ana Rita Serra Valente Caldeira

Dissertação para a obtenção do grau de Mestre em Ciências Biomédicas especialidade de Biologia Molecular em Medicina Tropical e Internacional

Janeiro 2017

Universidade Nova de Lisboa

Instituto de Higiene e Medicina Tropical

Propagation and Molecular Characterization of *Chlamydia trachomatis* strains isolated in Portugal

Autor: Ana Rita Serra Valente Caldeira

Título do Candidato: Licenciada em Anatomia Patológica, Citológica e Tanatológica

Orientadora: Doutora Maria José Borrego, Investigadora Auxiliar no Instituto Nacional de Saúde Doutor Ricardo Jorge

Coorientadora: Doutora Rita Castro, Professora Auxiliar no Instituto de Higiene e Medicina Tropical

Dissertação apresentada para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Ciências Biomédicas

Agradecimentos

Gostaria de agradecer:

À minha orientadora Dra. Maria José Borrego, Responsável pelo Laboratório Nacional de Referência das Infeções Sexualmente Transmissíveis do INSA, por me ter aceite no seu laboratório, por todos os conhecimentos transmitidos, por toda a paciência, pela permanente disponibilidade, amizade e apreço.

À minha coorientadora Dra. Rita Castro, Professora Auxiliar do IHMT, por ter aceite ser minha orientadora pela disponibilidade e simpatia.

Ao Dr. Celso Cunha, Coordenador do Mestrado de Ciências Biomédicas do IHMT, pela simpatia e disponibilidade.

Ao Dr. Vítor Borges, Investigador na Unidade de Bioinformática do Departamento de Doenças Infeciosas do INSA, pelo importante contributo dado à realização da presente tese, pelos conhecimentos científicos, simpatia e amizade.

Ao Dr. João Paulo Gomes, Responsável da Unidade de Investigação do Departamento de Doenças Infeciosas do INSA.

À Dora Cordeira, Técnica no Laboratório do Infeções Sexualmente Transmissíveis, pelos conhecimentos transmitidos e amizade desde o primeiro dia.

Aos demais colegas, nomeadamente: Miguel Pinto pela amizade, companheirismo e entreajuda; e, Alexandra Nunes, pela simpatia e entreajuda.

Agradeço também aos meus pais por todo o suporte para a realização do mestrado e por me dar a possibilidade de concluir mais uma etapa.

Por fim, agradeço ao Hugo Barreto por tudo o suporte, compreensão, paciência, disponibilidade e entreajuda. O principal pilar para a conclusão deste trabalho.

Resumo

Chlamydia trachomatis é uma bactéria intracelular que infeta exclusivamente o Homem. Constitui o principal agente etiológico bacteriano de infeções sexualmente transmissíveis e, na ausência de tratamento, pode evoluir para complicações clínicas graves, como a infertilidade tubária. C. trachomatis foi classificada em 15 genótipos, os quais causam preferencialmente infeções oculares (A-C), infeções genitais (D-K) ou linfogranuloma venéreo (LGV) (L1-L3). A tipagem das estirpes de clamídia baseia-se na principal proteína da membrana externa (MOMP) codificada pelo gene ompA. O estudo do genoma de C. trachomatis tem permitido identificar diferenças genéticas entre estirpes; no entanto, os mecanismos de adaptação e virulência não estão ainda bem definidos mas, tal tem sido observado noutras bactérias, podendo decorrer de mecanismos de variação de fase.

No presente estudo foi realizada a tipagem de 278 amostras clínicas com base na sequência do gene ompA, através de técnicas de nested-PCR, de sequenciação pelo método de Sanger e de análise de similaridades das sequências obtidas relativamente a sequências de estirpes protótipo de C. trachomatis; tal processo possibilitou identificar variantes genotípicas previamente descritas assim como novas variantes, sugerindo a existência de um contínuo processo de adaptação na bactéria. Os genótipos mais frequentes foram o E, D e F, tal como descrito na maioria dos estudos em populações heterossexuais femininas; contudo um número apreciável (26) de estirpes do genótipo L2 foi igualmente detetado, em particular em amostras anorretais masculinas, o que está igualmente de acordo com o descrito em populações de homens que têm sexo com homens. Tal constatação revela a necessidade de um método de identificação expedita dos genótipos LGV, pelo que contribuímos para a implementação de uma metodologia mais rápida que a tradicional genotipagem-ompA. Pretendemos ainda contribuir para a avaliação da virulência das estirpes de C. trachomatis e para tal, realizamos um estudo preliminar de identificação de potenciais alvos genéticos sujeitos a variação de fase, com recurso à sequenciação de nova geração (NGS).

Este estudo constitui um contributo para a vigilância epidemiológica das infeções por C. trachomatis em Portugal, uma situação pouco conhecida, um contributo para a celeridade da identificação de casos de LGV e um contributo para a avaliação da virulência das estirpes de C. trachomatis, pela identificação de alterações em genes potencialmente relacionados com a variação de fase.

Palavras-chave: Chlamydia trachomatis; genotipagem ompA; variação de fase.

Abstract

Chlamydia trachomatis is an intracellular bacterium that exclusively infects humans. It is the main bacterial etiological agent of sexually transmitted infections and, in the absence of treatment, may lead to serious clinical complications such as tubal infertility. C. trachomatis was classified into 15 genotypes, which preferentially cause ocular infections (A-C), genital infections (D-K) or lymphogranuloma venereum (LGV) (L1-L3). Typing of chlamydial strains is based on the major outer membrane protein (MOMP), which is encoded by the ompA gene. The study of the genome of C. trachomatis allowed identifying genetic differences between strains, and although the mechanisms of adaptation and virulence still require proper definition, these mechanisms may derive from mechanisms of phase variation, as has been observed in other bacteria.

In the present study, 278 clinical samples were typed based on their ompA gene sequence, determined by using a nested-PCR technique, Sanger method sequencing and gene sequence similarity analysis regarding the ompA-sequence of C. trachomatis prototype strains. This procedure allowed to identify genotypic variants previously described, as well as new variants, suggesting the existence of a continuous process of adaptation in the bacterium. The most frequent genotypes in our study were E, D and F; this finding is in accordance to the described in most of the studies held in female heterosexual populations. However, an appreciable number (26) of L2 strains was also detected, in particular in male anorectal samples, which is an in agreement with the described for men who have sex with men. The number of L2 strains reveals the need for an expeditious LGV-genotype identification method. Thus, we contributed to the implementation of a methodology that should provide faster identification of LGV strains in comparison to the traditional ompA-genotyping. We also intended to contribute to the evaluation of the virulence of C. trachomatis strains and for this, we participated in a preliminary study of identification of potential genetic targets subjected to phase variation, using a new generation sequencing (NGS) approach

This study contributes to the epidemiological surveillance of C. trachomatis infections in Portugal, a poorly known situation, contributes to a faster identification of LGV cases, and contributes to the evaluation of virulence of C. trachomatis strains by identifying changes in genes potentially related to phase variation.

Keywords: Chlamydia trachomatis; ompA-genotyping; phase variation.

Table of contents

Agradecimentos i
Resumo ii
Abstract iii
Table of contents iv
Figure index vi
Table index vii
List of abbreviations viii
I. Introduction 1
1. Sexually Transmitted Infections 1
2. Chlamydia 1
2.1. The *genus Chlamydia* 1
2.2 Chlamydia trachomatis 2
2.2.1 Biology: The life cycle of *C. trachomatis* 2
2.2.2 Epidemiology and Human diseases 6
2.2.3 Genomics features 11
3. Objectives 13
II. Material and Methods 15
1. Cell culture and inoculation of *C. trachomatis* 15
2. Typing *C. trachomatis* and bioinformatic analysis 19
3. Strategies for quick subtyping LGV strains 23
3.1 Subtyping LGV: Using sets of primers discriminatory of L1, L2 and L3 23
3.2 Subtyping LGV: real time-PCR for *C. trachomatis ompA* and *pmpH* 24
4. Homopolymeric tracts counts and bioinformatics analyses 26
4.1 Study design: selection of homopolymeric tracts and *C. trachomatis* positive DNA samples 26
4.2 Amplicon-based NGS and bioinformatics analyses 27
III. Results and Discussion 31
1. Cell culture and *C. trachomatis* typing 31
1.1 Cell culture, the old gold standard 31
1.2 Characterization of *C. trachomatis* clinical strains 33
2. Strategies for quick subtyping LGV strains 44
3. Phase variation in *C. trachomatis* – Preliminary study 47
3.1 Validation of the approach 47

3.2 Allelic variation among strains 51
IV. Conclusion 59
V. References 61
VI. Annexes 75

Figure index

Figure 1 - Schematic representation of the development cycle of *C. trachomatis* 3

Figure 2 – Schematic representation of *C. trachomatis* outer membrane 5

Figure 3 – Phylogenetic tree of *C. trachomatis* 7

Figure 4 – Schematic representation of inoculation in a 24 well plate 18

Figure 5 – Schematic representation of homopolymeric traits counts 29

Figure 6 – Results from propagating *C. trachomatis* strains in cell culture 31

Figure 7 – Rate of *C. trachomatis ompA*-genotyped strains 33

Figure 8 – *C. trachomatis* strains by gender 34

Figure 9 – *C. trachomatis* strains by age 35

Figure 10 – Frequency of biological samples positive for *C. trachomatis* 36

Figure 11 – Frequency of the different *ompA*-genotypes among the studied *C. trachomatis* strains 37

Figure 12 – Number of *ompA* variants among detected *C. trachomatis* genotypes 39

Figure 13 – *C. trachomatis ompA*-genotype distribution according to anatomical site of infection 42

Figure 14 – *C. trachomatis ompA*-genotype frequency per age group 43

Figure 15 – Conventional PCR approach for detecting L1, L2 and L3 45

Figure 16 – Results for real time PCR assay 46

Figure 17 – Comparison between WGS and amplicon-based NGS techniques 49

Figure 18 – Comparison between 35 and 45 cycles PCR 50

Figure 19 – Profile of the CT166 homopolymeric tract in several *C. trachomatis* strains 52

Figure 20 – Profile of the homopolymeric tract affecting the regulatory region of CT533 for several strains 53

Figure 21 – Analysis of tropism-specific homopolymeric lengths 55

Figure 22 – Intra-strain homopolymeric tracts variability 57

Table index

Table 1 – Formula for complete medium preparation 16

Table 2 – Formula for enriched medium 17

Table 3 – Nested-PCR mix 20

Table 4 – *C. trachomatis ompA* Nested-PCR amplification profile 20

Table 5 – Sequencing mix 21

Table 6 – Sequencing profile 22

Table 7 – Master mix for LGV subtyping real time – PCR 24

Table 8 – Amplification profile for LGV subtyping real time – PCR 25

Table 9 – PCR mix for homopolymeric targets 27

Table 10– Amplification profile for homopolymeric regions 28

Table 11 - Nucleotide sequence variation within *ompA* in relation with the respective prototype strain 40

List of abbreviations

µg – Microgram

µl - Microliter

µm – Micrometer

2SP – 2 sucrose phosphate

ATCC – American Type Culture Collection

bp – Base pair

CD – Conserved domain

CO2 – Carbon dioxide

COMC – Chlamydia Outer Membrane Complex

ddNTPs – Dideoxynucleotide

DFA – Direct Fluorescence Antibody

DMSO – Dimethyl Sulfoxide

DNA – Deoxyribonucleic acid

dNTP – Deoxynucleotide

DPBS – Dulbecco's Phosphate Buffer Saline

EB – Elementary Body

EDTA – Ethylenediaminetetraacid

EIA – Enzyme Immunoassays

ELISA – Enzyme-Linked Immunosorbent Assay

FBS – Fetal Bovine Serum

HBV – Hepatitis B Virus

HCV – Hepatitis C Virus

HIV – Human Immunodeficiency Virus

HPV – Human Papilloma Virus

HSV – Herpes Simplex Virus

IFN-γ – Interferon-γ

IgA – Immunoglobulin A

IgG – Immunoglobulin G

IgM – Immunoglobulin M

INSA – National Institute of Health (from the Portuguese Instituto Nacional de Saúde)

kb – Kilobase (= 1 000 bp)

LGV – Lymphogranuloma venereum

LPS – Lipopolysaccharide

MACPF – Membrane Attack Complex/ Perforin

Mb – Megabase (= 1 000 000 bp)

MEM – Minimum Essential Medium

mg – Milligrams

MgCl2 – Magnesium chloride

ml – Milliliter

mM – Millimolar

MOMP – Major Outer Membrane Protein

MSM – Men who have Sex with Men

NAATs – Nucleic Acid Amplification Tests

NGS – New Generation Sequencing

ºC – Degrees Celsius

ORF – Open Reading Frame

PCR – Polymerase Chain Reaction

PID – Pelvic Inflammatory Disease

PLD – Phospholipase D

pmol – Picomolar

PMP – Polymorphic Membrane Proteins

POC – Point-Of-Care

RB – Reticulate Body

RFLP – Restriction Fragment Length Polymorphism

STI – Sexually Transmitted Diseases

T3SS – Type III Secretion System

TAE – Tris-Acetate-EDTA

TARP – Translocated Actin Recruiting Phosphoprotein

UTI – Unit of Technology and Information

v - Volume

V – Volts

VD – Variable Domain

w – Weight

WGS – Whole Genome Sequencing

WHO – World Health Organization

I. Introduction

1. Sexually Transmitted Infections

Sexually transmitted infections (STI) can be caused by different microorganisms, such as: a) bacteria (like *Chlamydia trachomatis*, *Neisseria gonorrhoeae*, *Treponema pallidum*, *Haemophilus ducreyi*, *Mycoplasma genitalium* and *Ureaplasma urealyticum*), b) virus (Human Immunodeficiency Virus [HIV], Herpes Simplex Virus [HSV] type 1 and 2, Human Papilloma Virus [HPV], Hepatitis B Virus [HBV] and Hepatitis C Virus [HCV]), c) parasites, either protozoa (*Trichomonas vaginalis*) or metazoa (*Phthirus pubis* and *Sarcoptes scabiei*) (1).

In 2016, WHO (2), estimated in 357 million the number of new cases of the four curable STI, being *C. trachomatis* the second most frequent with 131 million after *T. vaginallis*, next *N. gonorrhoeae* and *T. pallidum*. There are several factors associated to mortality and morbidity of STI, such as fetal and neonatal death, cervical cancer, infertility and higher risk of HIV acquisition.

2. Chlamydia

2.1. The *genus Chlamydia*

C. trachomatis is an obligate intracellular bacterium and one of the eleven species within the family *Chlamideaceae*, order *Chlamydiales*, phylum *Chlamydiae* and genus *Chlamydia*. Beyond *C. trachomatis*, the *Chlamydia* genus also includes *C. pneumoniae, C. pecorum*, *C. felis*, *C. psittaci*, *C. abortus*, *C. caviae*, *C. avium*, *C. gallinacea*, *C. suis* and *C. muridarum*. All these species differ in what concern natural host (animal or

human) and tissue tropism (3–5). The most relevant chlamydial species infecting humans are *C. trachomatis*, that only infects humans and is associated with ocular and anogenital infections, the later, sexually transmitted (6) and *C. pneumoniae* (7), related to a wide range of diseases including respiratory illnesses, arthritis (8) and some chronic pathologies involving the cardiovascular (9), or the central nervous systems (10). *C. psittaci* can also infect humans, who usually are accidental hosts, among who some genotypes are more prone for causing disease, called psittacosis, a pathology that affects the respiratory system (11).

2.2 Chlamydia trachomatis

2.2.1 Biology: The life cycle of *C. trachomatis*

The family *Chlamideaceae* is constitute by obligate intracellular bacteria that evolved by acquiring the ability to adapt and colonize its host. Therefore, *C. trachomatis* presents a unique and complex biphasic development cycle along which it alternates between two morphologically distinct forms: the elementary body (EB) capable of extracellular survival for a short period, infectious and metabolically inactive and the intracellular reticulate body (RB), non-infectious and metabolically active (12).

The EB's are small gram-negative cocci (~ 0,3µm) presenting a polarized architecture while RB's are larger (~ 1µm) (13).

The life cycle comprehends four major events: it initiate with binding and invasion of the host cell (14), that is followed by differentiation of EB into RB that undergo several cell division cycles before re-differentiate into EB that are liberated to infect new cells by the end of the cycle upon eclosion of the inclusion/ host cell (13). Until 2016, chlamydial cell division was believed to occur through binary fission, despite the absence of FtsZ protein in these bacteria; however, this year, it has been shown that *Chlamydia* divides by a polarized cell division process through which major outer membrane protein (MOMP) first restricts to one pole of the bacterial cell, where the chlamydial cell division

machinery also accumulates, and it is from there that the bacterial cell daughter emerges, a so called budding process (15). A complete development cycle takes 48 to 72 hours and can be interrupted by environmental stress factors, like nutrient deprivation or exposure to antibiotics (12,16). (Figure 1).

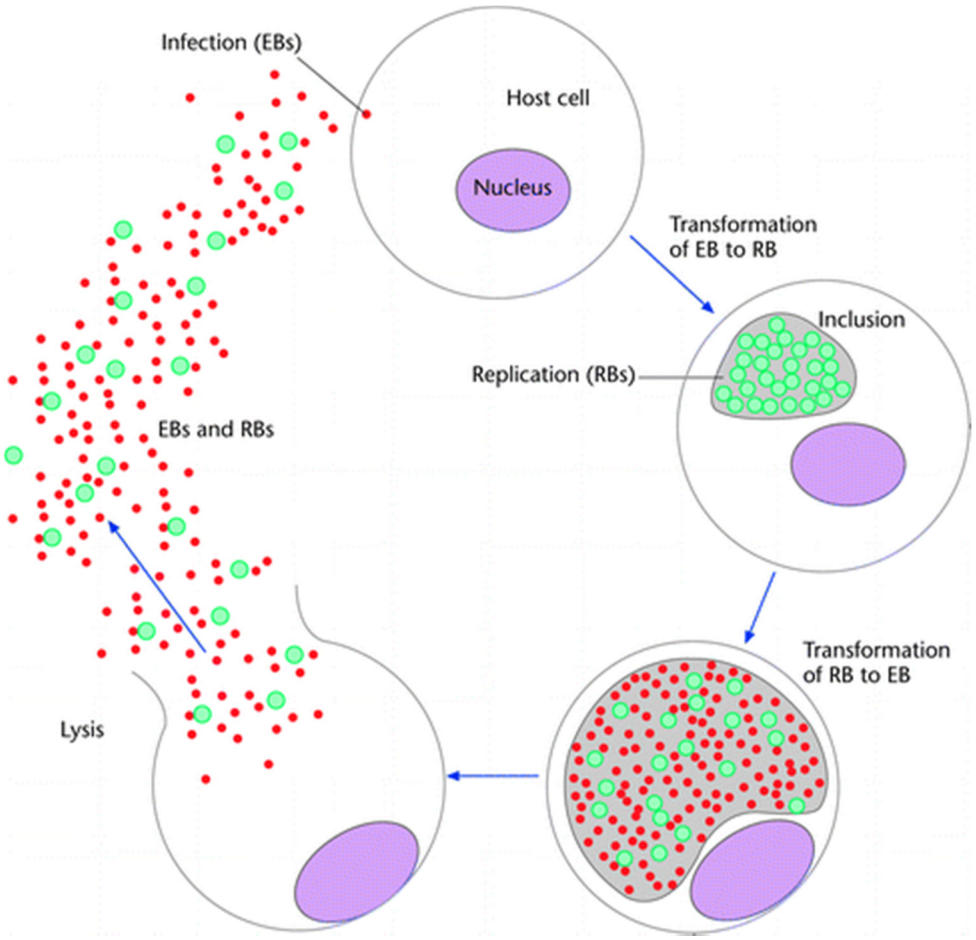

Figure 1 - Schematic representation of the development cycle of *C. trachomatis.* [Available in (17)]

C. trachomatis, as any gram-negative bacteria, is provided of a double membrane; the outer membrane exhibits a type III secretion system (T3SS), the chlamydia outer membrane complex (COMC), polymorphic membrane proteins (Pmp), porins, lipids and polysaccharides (18,19) (Figure 2).

The T3SS has been identified in *C. trachomatis* in 1997 (20), firstly described as rosette structure present in the outer membrane. A set of T3SS is located in one of the EB hemisphere (14) and composes a macromolecular structure called "injectisomes" constitute by several proteins (21). The T3SS is a widely used secretion system among gram-negative bacteria and is essential for virulence of human pathogens (22). For chlamydial species it is essential for bacterial protein delivery across the eukaryotic cell membrane, contributing to chlamydial pathogenesis (21,23).

The COMC is constitute by the MOMP and by two other major proteins, the OmcA and the OmcB (24–26). MOMP represents 60% of membrane dry weight (25), and has been associated to pathogenesis given its variability and antigenic properties (27). MOMP also acts as a cytoadhesin (28), interacting with the host, and has been characterized also as a porin (29). Beyond MOMP another porin has been identified in the outer membrane, the PorB, a conserved porin that seems to function exclusively for transport of nutrients for chlamydial survival (30).

Also present in the outer membrane are the pmp, which were characterized as autotransporters and adhesins, being coded by a highly heterogenic gene family presenting a different number of *pmp* genes amongst the different chlamydial species (31,32).

Well represented in the surface of *C. trachomatis* is the lipopolysaccharide (LPS) being the main lipid present in the outer membrane (33). LPS functions as a barrier to permeability (19) enabling cell viability, and beyond that, it has been related to interactions with host cells, and thus associated with infectivity (34).

In 2013 (35) it was first demonstrate the presence of peptidoglycan between the double membrane of the *Chlamydia* cell wall; its functions were related to cellular integrity and it should also play a role during cell division. The goal of detecting peptidoglycan in the chlamydial cell wall was not an easy task; however, recent labeling and proteomics methodologies made that possible (35,36).

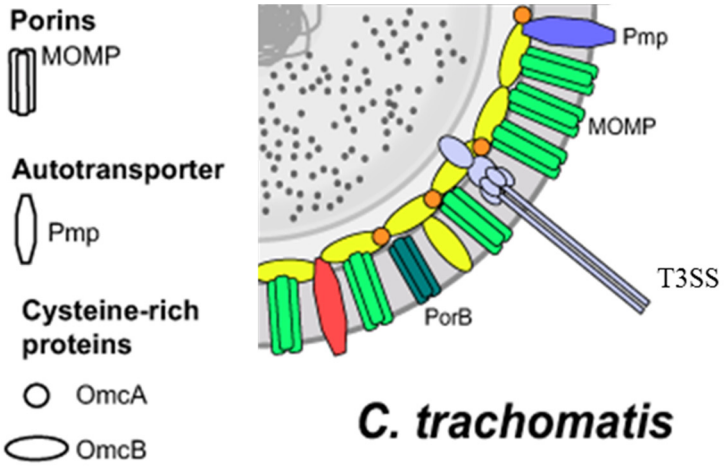

Figure 2 – Schematic representation of *C. trachomatis* outer membrane [Adapted from (37)]

Since EBs start interact with the host cell, bacterial proteins entry in the cytoplasm of the host cell mediated by the T3SS (22). Next, several effectors are recruited to entry location; one of them is the translocated actin recruiting phosphoprotein (TARP) that will be involved in the inclusion formation. The chlamydial inclusion is a cytoplasmic vacuole surrounded by host cell membrane, where specific proteins called Inc, are identifiable. Incs interact with the host cell during infection, becoming important factors for the maintenance and growth of *C. trachomatis* (38,39).

Gene transcription starts when EB differentiate into RB. Genes have been categorized according to three temporal transcription classes (in relation to the chlamydial life cycle) in: early, midcycle and late (13,40). Early genes are related to inclusion remodeling and host-pathogen interaction. Midcycle genes, which are the largest group, are associated with nutrient acquisition, growth and replication. Finally, late genes start to be transcribed after the second differentiation (RB to EB) and are related to DNA condensation and the outer membrane complex formation (13,18). To note, the second differentiation occurs in an asynchronous manner, which means that the differentiation of RBs into EBs do not occur simultaneously in all particles (12,13).

There are two critical phases during chlamydial infections, the entry in the host cell to perform the infections and the exit of the host cell to keep the infection. The exit from the host cell usually happens by a lytic mechanism, through which a new progeny

of EBs is released (13), after inclusion and host cell membranes rupture. The lytic mechanism has been associated to LGV infections and it is thought to be regulated by the plasmid (41). A non-lytic exit is characterized by the extrusion of the whole inclusion and involves host cell actin and myosin allowing chlamydial survival inside macrophages and infection at further anatomic sites (42–44).

2.2.2 Epidemiology and Human diseases

C. trachomatis is the most common bacterial STI, with an estimated incidence of 131 million new cases per year, affecting in particular individuals aged 15 to 49 years (2). Moreover, *C. trachomatis* infections potentiate HIV transmission due to epithelial damage (45).

Three biovars can be identified in *C. trachomatis*, related to tissue tropism. These biovars are named trachoma, genital tract and lymphogranuloma venereum (LGV) (13). Each biovar includes several bacterial types that were first identified through the use of monoclonal antibodies developed against *C. trachomatis* MOMP (46). The nucleotide differences of the MOMP encoding gene, *ompA*, reflect the differences found in the amino acid sequence and for this reason the 15 genotypes are defined reflecting the 15 serovars (A, B, Ba, C, D, E, F, G, H, I, J, K, L1, L2 and L3) (47). The majority of the amino acid sequence of MOMP is conserved, except within four variable domains (VD). These VD are responsible for the major antigenic differences among the 15 identified *ompA*-genotypes (48,49) (Figure 3).

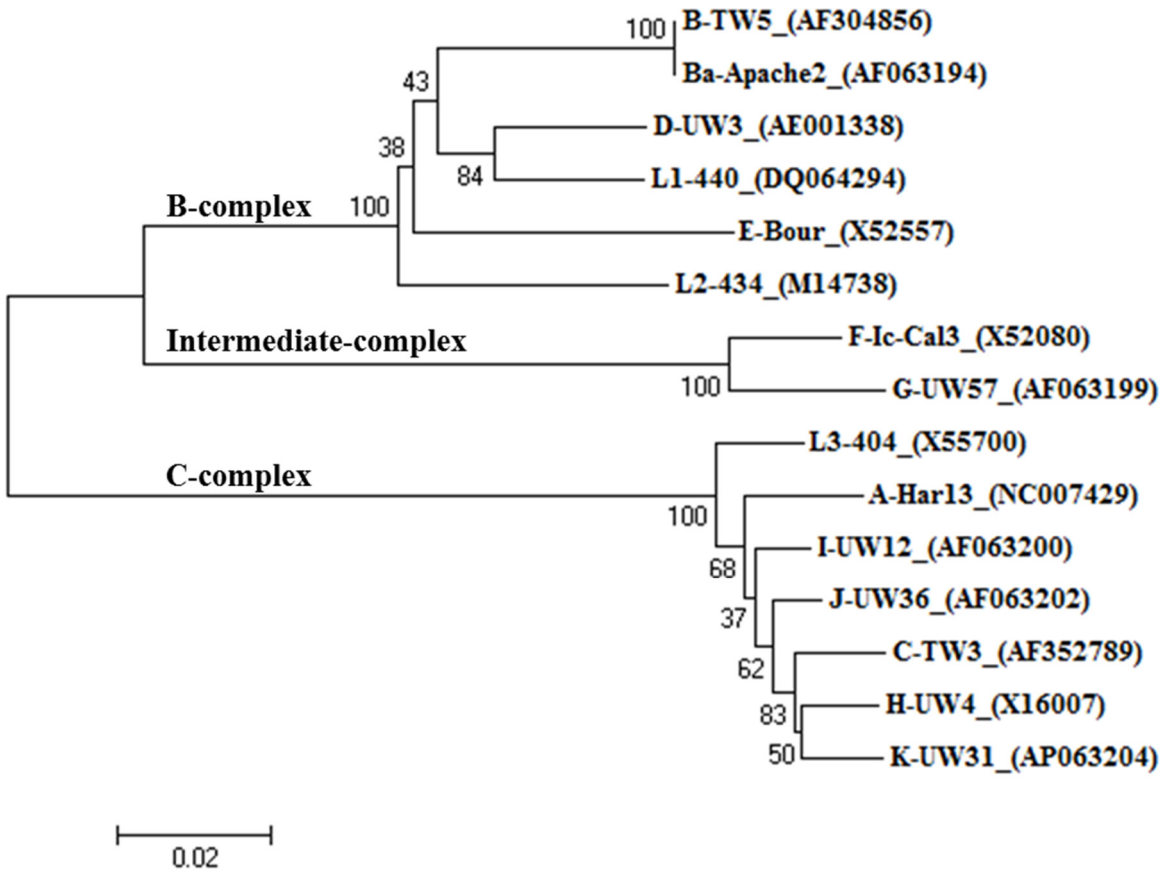

Figure 3 – Phylogenetic tree of *C. trachomatis*. The neighbor joining phylogenetic tree (50) is based on the nucleotide differences among *ompA* sequences, computed using the Kimura 2-parameter method (51). Bootstraps values (1000 replicates) are shown next to the branch nodes. Evolutionary analyses were conducted in MEGA7 (52).

Based on *ompA* gene diversity, three phylogenetic groups can be defined: the B-complex (includes *ompA*-genotypes B, Ba, D, E, L1 and L2), the C-complex (includes *ompA*-genotypes A, C, H, I, J, K and L3) and the intermediate complex (includes *ompA*-genotypes F and G) (48,53). These three phylogenetic do not reflect the three biovars, indicating that chlamydial *ompA* genotyping does not reflect the whole genome in terms of pathobiology (53).

In terms of associated pathology, *ompA*-genotypes A to C cause trachoma while *ompA*-genotypes D to K constitute the major bacterial cause of STI and *ompA*-genotypes L1 to L3 cause LGV (13).

Trachoma affects mainly the conjunctival epithelia, being the most common infectious cause of blindness, yet preventable, related to poverty and poor hygiene conditions (54). Trachoma affects 40 million people, being endemic in 50 countries and includes the list of 17 neglected tropical diseases (55).

C. trachomatis D-K *ompA*-genotypes are the major cause of bacterial STI worldwide (6). *C. trachomatis* genital infection is asymptomatic in about 70% of women and 50% of men (56). When existing, the clinical manifestations in males include urethritis, epididymitis, epididymo-orchitis, prostatitis and infertility (57,58). In women *C. trachomatis* genital infections, besides the clinical manifestations as cervicitis and urethritis, may also evolve to complications such as salpingitis, endometritis, ectopic pregnancy, pelvic inflammatory disease (PID) and tubal infertility (6,59). Associated to the genital infections, *C. trachomatis* can also be transmitted to newborns during delivery. These neonatal infections appear as conjunctival or nasopharyngeal infection that can progress to pneumonia (6,60).

LGV *ompA*-genotypes beyond the capacity of infecting epithelial cells (like other *ompA*-genotypes), can also infect mononuclear phagocytes, an invasive characteristic, as migration through immune cells allow LGV genotypes to infect regional lymph nodes or lymphatic tissue of the intestine (61). The clinical manifestation of LGV are diverse; the classic clinical picture included the appearance of a genital ulcer, followed by an inguinal stage during which some patients presented the "groove sign" that was characterized by the swallow of the inguinal nodes, finally a third stage often called "anogenitorectal syndrome", described as a chronic inflammatory lesion that leads to lymphatic obstruction (61–64). Endemic in several tropical countries such in some parts of Africa, Asia, South America and the Caribbean, LGV has been spreading since 2004 to industrialized countries (61,65). Since then, several outbreaks have been reported (66), such as in France (67), Great Britain (68), The Netherlands (69), Canada (70), Australia (63) and Portugal (71,72). L2 became the most frequent *ompA*-genotype associated to the recent LGV epidemics (65,66), during which new variants have been discovered, namely one L2b subtype associated to the Sweden outbreak (73). Curiously, this new Swedish L2b variant presented a deletion of 377 base pairs within the plasmid target region of the most popular molecular diagnostic method in Sweden, as discovered through gene sequencing (73); consequently, false negative cases escaped from treatment and were

allowed to propagate the infection. The diagnostic of *C. trachomatis* through acid nucleic amplification had become the gold standard in terms of diagnostic in the 2000s, and the omnipresent 4 – 8 copies *C. trachomatis* plasmid appeared as an attractive target (74,75). However, as a consequence of the Swedish outbreak, commercial nucleic acid amplification tests implemented multi-gene target systems, for avoiding false negatives caused by mutation, insertion or deletion phenomena that can affect the performance of single gene target based tests. It also contributed to show the importance of epidemiological data and external quality assessment programs (74).

Worldwide several molecular epidemiology studies have been done, evidencing that the most common *ompA*-genotype among heterosexuals are E, D and F, while G, D, J and L2 are most frequent among MSM (76).

In terms of public health, the major problem related to *C. trachomatis* STI is the high number of asymptomatic cases that remain untreated and continue propagating, resulting in millions of new infections per year. When left untreated, *C. trachomatis* infections may progress to severe reproductive tract complications and sequelae, like PID, ectopic pregnancy and tubal infertility (6,77). At present, the recommendations for treating *C. trachomatis* are azithromycin and doxycycline (77,78). Azithromycin is recommended for uncomplicated genital chlamydial infections in a single dose of 1 gram, while 100 milligram of doxycycline twice a day during 7 or 21 days is prescribed for anorectal chlamydial infections and LGV cases, respectively (78).

Due to the high number of asymptomatic cases, the guidelines for the management of chlamydial infections include the screening for STI in high risk populations (such as sexworkers), partner notification and management of pregnant woman in order to prevent adverse clinical outcomes for the newborn (1,79,80).

The serological techniques were among the earliest diagnostic methods for *C. trachomatis*, determining the antibody response to chlamydial antigens (81,82). Microimmunofluorescence is the recommended serological technique, although several enzyme-linked immunosorbent assay (ELISA) tests (6) are commercially available for detecting specific IgM, IgG an IgA. However cross reactivity to different *Chlamydia* species is common and antibody responses can be scarce; thus, serological testing should

be restricted to diagnose newborn infections, and to be used as a marker in case of LGV suspicion or tubal infertility (81–83).

To achieve a better management of chlamydial infections, detection techniques have greatly evolved since the isolation in cell culture, being the first isolation performed in 1957 (84). Ever since, the isolation of chlamydial strains through cell culture has been considered as a gold standard; however, time consumption and stringent transport (Sucrose Phosphate buffer – 2SP) and storage conditions (4ºC up to 24 hours, -80ºC for longer periods), together with the delay to obtain a result, have limited its applicability (82,85). For these reasons, chlamydial culture is nowadays restricted to reference laboratories and research (82).

Antigen detection based diagnostic techniques have also been developed, namely enzyme immunoassays (EIA) and the direct fluorescence antibody (DFA) (6). The EIA techniques were also implemented for use as point-of-care (POC) tests to allow diagnosis during a clinical visit; however, all EIA tests reveal low sensitivity and poor specificity, and for this reason their use is no longer recommended (82). DFA performance relies on monoclonal antibody quality, on fluorescence microscopy availability and most of all, on personal skills for recognizing EBs, a process that is time consuming; furthermore, antigen detection techniques are restricted to exudates (82).

The major advance of chlamydial infections diagnosis was the introduction of nucleic acid amplification tests (NAATs), which are nowadays the recommended due to their superior performance characteristics in terms of specificity and sensitivity, close to 100% (81,82). Moreover, these tests allow the use of first void urine, a non-invasive biological sample that is a well-accepted and recommended sample for male urethra evaluation. Several NAATs methodologies have been developed, but real-time polymerase chain reaction (PCR), is the most used. In addition, recent NAATs allow the simultaneous amplification and detection of various STI agents, in particular the two most common bacterial STI, *C. trachomatis* and *N. gonorrhoeae* (82). Some NAATs include several targets for each microorganism to avoid false negatives caused by target gene changing's (as explained above); as an example, COBAS 4800® has two targets for *C. trachomatis*, one in the chromosome and the other in plasmid (86).

2.2.3 Genomics features

The genomes of intracellular parasites are relatively small when compared to genomes of free-living microorganisms, due to the adaptation to their lifestyle (4,87). In fact, intracellular bacteria suffered a genome reduction with genetic material loss while becoming metabolic parasites, through which the bacteria benefits from living within the host cell cytoplasm (87,88).

Accordingly, the *Chlamydiaceae* family presents a small genome (87), specifically, *C. trachomatis* has a single circular chromosome with approximately 1Mb, where around 935 genes were identified, 887 of them are coding genes (89–91) and a 7,5 kb plasmid (only a few plasmidless strains were described), comprehending 8 coding genes (89,90,92). Excluding the plasmid, there are no evidence of other extrachromosomal or mobile genetic elements (90,91).

Genomic analysis of *C. trachomatis* shows a highly conserved genome with the exception of a 50 kb region of high variability, named plasticity zone, dissimilar among chlamydial strains (90). This hypervariable region contains some genes related to pathogenesis and to interactions with the host. Some of the genes present in the plasticity zone are the tryptophan synthase (*trp*) operon, the cytotoxin, the membrane attack complex/ perforin (MACPF) and phospholipase D (PLD) (91). The *trp* operon is considered a virulence factor, as it provides the ability to synthesize tryptophan in the presence of interferon (IFN)-γ from the host immune response (93). However this *trp* only exists in genital *ompA*-genotypes, which are the only having a functional tryptophan synthase. The chlamydial cytotoxin, a highly polymorphic gene, appears differently among strains, but only genital *ompA*-genotypes encode both functional domains of the cytotoxin ORF (CT166), the glycosyltransferase and UDP-glucose binding domains. Ocular *ompA*-genotypes encode only one domain, the UDP-glucose binding domain, and LGV *ompA*-genotypes lack the CT166 ORF. The chlamydial cytotoxin has a significant homology with the clostridial cytotoxin, and leads to a cytopathic effect since it is involved in disabling the actin filaments of the host cell cytoskeleton (94–96). Other genes, like the MACPF and the PDL that were also related to pathogenesis, are present in all *ompA*-genotypes; their functions have been further related to acquisition of

metabolites from the host, a function that is essential for the chlamydial survival and successful development cycle (97).

Outside the plasticity zone, other genes have been related to virulence and pathogenicity such as *Pmp* genes, TARP genes, *Inc* genes and the *ompA* gene (98). Genes encoding the Pmp family, unique of the order *Chlamydiales*, constituted by 9 members (Pmp A to I) localized in the outer membrane, involve 3,2% of the genome coding capacity. Pmps are also described as autotransporters, pointed as virulence factors (because they promote antigenic polymorphisms), adhesion and differential tissue tropism (31). The type III secretion system effector TARP is involved in actin remodeling during invasion; it exhibits a high degree of variability and it is known to be involved in niche adaptation (99). The analysis of the TARP of different *ompA*-genotypes showed variation in the number of tyrosine-rich repeat regions and in the actin binding domain. In respect to tyrosine-rich repeat regions the ocular strains have the lowest number, while LGV strains present the highest; related to acting binding domains, the ocular strains have a higher number than LGV strains; genital strains are in the middle for both (100). Inc's are also effectors of the type III secretion system family and are proteins inserted in the inclusion membrane and thus exposed to the host cytosol. Inc proteins are species-specific, and very conserved among *ompA*-genotypes, although evidencing some divergence between strains with different tissue tropism (39). MOMP, encoded by the *ompA* gene, is a porin implicated in attachment (functioning as an adhesin), in pathogenesis, due to its variability and antigenic properties (includes genus, species, and type epitopes and induces humoral and cellular immune response in the host, justifying being considered a potential candidate for the development of a vaccine (27)) and, recently, it was further implicated in cell division (15).

Besides the knowledge of genetic variation between strains, phase variation mechanisms have been shown to play a role in adaptation and pathogenesis in *C. pneumoniae* (101,102) and it could be expected that the same functions occur in *C. trachomatis* (32,103–105). The phase variation mechanism relies on the ability of bacteria to quickly adapt in response to a stimuli, being usually associated with a reversible switching between an ON and OFF state of specific proteins, such as: surface-exposed proteins or proteins involved in the biosynthesis/ expression of the bacterial capsule and the LPS (a major component of gram-negative bacteria surface); the switching occurs due

to changes in DNA that lead to different phenotypes (106). The phase variation mechanism in many cases occur due to expansion and contraction of homopolymeric tracts, described as repetitive regions of a DNA single base [poly(A), poly (T), poly (G) and poly (C)] (107,108).

3. Objectives

- Propagation of *C. trachomatis* clinical strains in cell culture
- *ompA*-genotyping of *C. trachomatis* clinical strains
- Implementation of molecular biology techniques to specifically and rapidly detect L1, L2 and L3 *ompA*-genotypes
- Phase variation evaluation through allele homopolymeric tract counting

II. Material and Methods

1. Cell culture and inoculation of *C. trachomatis*

During our work we used the HeLa 299 immortal cell line (American Type Culture Collection® [ATCC] CCL – 2.1™), which derives from cervical cancer cells, and is commonly used for the propagation of *C. trachomatis*.

The specimens inoculated in cell culture were cervical, urethral and anorectal exudates kept in 2SP buffer at -80ºC upon arrival to the laboratory. These specimens had tested positive for *C. trachomatis* in the host laboratory during its routine diagnosis service, through COBAS 4800® (Roche). INSA provides routine *C. trachomatis* diagnosis to attendees of STI, family planning, general practice, urology and gynecology clinics, among others, and also participates in *C. trachomatis* screening studies in specific populations (i.e. sex workers). *C. trachomatis* prototype strains were also inoculated, namely G-UW57 (ATCC® VR-878™), H-UW43 (ATCC® VR-879D™), I-UW12 (ATCC® VR-880™), F-IcCal3 (ATCC® VR-346™), and L3-404 (ATCC® VR-903™).

For culturing HeLa 299 cells we used the Minimum Essential Medium (MEM) (41090 Gibco™) supplemented with fetal bovine serum (FBS) (16000 Gibco™), antibiotic and antifungal agents, hereby designated by "complete medium" (Formula in Table 1). In a first step and using a water bath at 37ºC, we defrosted an aliquot of cells stored at -195,79ºC in liquid nitrogen. After defrosting, cells were put in a T25 flask in which 8 ml of complete medium were added together with an additionally 10% of FBS. Cells were left to incubate at 37ºC, 5% CO_2 for 24 hours to obtain a cell monolayer that could be observed with an inverted microscope (objective x20). The maintenance of the cell culture is performed by passaging, a process that involves two times washing with 10 ml of Dulbecco's Phosphate-Buffered Saline (DPBS) (14190 Gibco™), for medium and dead cells removal, cell detaching by adding 4 ml of 1x trypsin solution (15090 Gibco™), suspension of cells with fresh complete medium (10 ml) and transfer of a variable volume

(depending on time until next passage and cell culture surface area) to a new tissue culture flask containing fresh complete media.

Table 1 – Formula for complete medium preparation

Supplements (in %) to complete MEM (41090 Gibco™)	
0,3%	Gentamicin (10 mg/ml)*
0,3%	Amphotericin B (250 µg/ml)**
11%	Fetal Bovine Serum (FBS)***

*15710 **15290 ***16000 Gibco™

We also prepared HeLa 299 cell stock, by adding 1 ml of dimethyl sulfoxide (DMSO, Amresco®) to a 10 ml cell-suspension (10% v/v) from a T175-flask confluent monolayer, distributing by 10 cryotubes (1 ml/cryotube) and allowing a slow drop of the temperature until -80ºC by using a "Cryo 1ºC freezing container" (Nalgene™). Frozen aliquots of HeLa 229 cells were then stored in a liquid nitrogen container (or at -80ºC for a shorter usage).

For inoculation of *C. trachomatis* positive specimens or prototype strains, we transferred HeLa 299 cells into a 24 well plate at a final concentration of $2{,}3x10^5$ cells/ml/well (see Annex 1) and let to incubate for 24 hours at 37ºC, 5% CO_2. For ATCC strains we removed 0,8 ml of medium from each well and added 0,2 ml of prototype strain suspension. For clinical samples, we removed 0,5 ml of medium from each one of four wells and added 0,2 ml of the sample to each one of four wells. The inoculated plates were centrifuged 1 hour at 32ºC, 2126 g in order to promote adhesion of *C. trachomatis* to the cell monolayer and let to incubate for another hour at 37ºC, 5% CO_2. The supernatant was replaced by "enriched medium" (Formula in Table 2), followed by incubation at 37ºC, 5% CO_2 for 48 hours.

Table 2 – Formula for enriched medium

Supplements (in %) to prepare enriched medium	
1%	Vitamins (100x)*
1%	Non-essential aminoacids (100x)**
5,4%	Glucose (10x)[a]
0,5%	Cycloheximide (100 µg/ml)[b]
Until perform the total volume	Complete medium

* 11120 ** 11140 Gibco™ [a] INSA [b] Sigma-Aldrich

C. trachomatis characteristic inclusions can be observed under an inverted microscope (objectives x10 and x40) after 48 hours of incubation. However, inclusions are better observed and counted by fluorescence microscopy (objective x20 and x40). For this, when preparing the cell culture in the 24 well – plate, a 12 mm diameter glass coverslip was placed at the bottom of one of the selected four wells for each specimen (Figure 4). After the 48 hours incubation, media from coverslip containing wells is withdrawn, 1 ml of methanol is added for fixing and coverslips are detached for further staining with a monoclonal antibody specific for *C. trachomatis* (Pathfinder *Chlamydia trachomatis* Direct Specimen Monoclonal Antibody kit) (Bio-Rad®), according to the manufacturer's instructions.

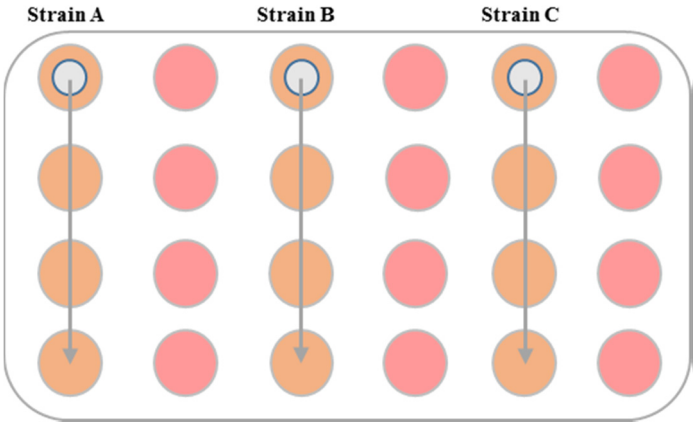

Figure 4 – Schematic representation of inoculation in a 24 well plate. Schematic representation of a 24 well-plate, where in pink are represented non-inoculated wells and in orange are represent inoculated wells; in the latter, glass coverslips have been prior added to the first well of each column.

For negative and low efficiency inoculations (1 – 4 inclusions per microscope field) we scraped the cell monolayer, collected the whole content of the well, performed cell disruption through sonication (Bioblock scientific, vibracell®) for 7 minutes, 80% Cycle Actif, power 9 and briefly centrifuged at room temperature for 7 minutes at 99 g, to eliminate cell debris, prior repeating inoculation procedure (keeping a 500 µl aliquot + 500 µl 2SP at -80ºC as a stock). When a high inoculation efficiency (≥ 5 inclusions per microscope field) is observed, cultures were treated as described above but they did not require further inoculation; thus, supernatant was divided into two cryotubes and added with an equal volume of 2SP, one to be kept at -80ºC and another to be stored in liquid nitrogen.

We also prepared stocks of chlamydial prototype strains. The preparation of these stocks involved the inoculation of aliquots of prototype strains (stored at -80ºC or liquid nitrogen) on HeLa 229 monolayers prepared in 24 well plates. Chlamydial cultures treated as described above and when culture efficiency reached ≥ 5 inclusions per microscope filed, supernatants were inoculated into HeLa 229 monolayer prepared in T25 flasks. Prototype chlamydial culture was controlled by inverted microscopy (objective x10 and x40), and when culture efficiency again reached ≥ 5 inclusions per microscope field and just prior inclusion rupture by the end of the chlamydial life cycle, cell monolayer was detached with glass beads and 4 ml of complete medium. This chlamydial suspension was sonicated, further centrifuged (as described above), and distributed by cryotubes to which

an equal volume of 2SP buffer is added. We performed a slow mixing of the cryotubes prior to freezing in "Cryo 1 ºC freezing container" (Nalgene™) and further storing at -80ºC or liquid nitrogen.

2. Typing *C. trachomatis* and bioinformatic analysis

C. trachomatis typing was performed over samples that: a) arrived at INSA for diagnosis and revealed positive; b) in specimens that were determined positive in other laboratories and sent to INSA for national collection purposes. The specimens were exudates (cervical, urethral, anorectal and conjunctival) and urines. For exudates, 200µl of the original sample were taken for DNA extraction, while for urines, DNA was extracted from 1 ml. DNAs were kept at -20ºC until further investigation.

C. trachomatis DNA was extracted using the NucliSENS®easyMag® of bioMérieux, according to manufacturer's instructions and using the specific EasyMag equipment.

A nested-PCR technique was used for *ompA* typing, through which two sequential PCR amplifications are performed in order to increase sensitivity. Primers used for typing are listed in Annex 2. In the first amplification (outer amplification) we used the primers NLO and NRO, and for the second amplification (inner amplification) we used the primers PCTM3 and SERO2A (109). We prepared PCR mixes (Table 3) for a total volume of 15 µl that were added with 10 µl of each DNA sample for the first PCR; for the second PCR, we prepared a total mix volume of 23 µl that were added with 2 µl of PCR product from the first round of amplification. The two pairs of primers are specific of the *C. trachomatis ompA* gene and promote an amplicon of about 1000 bp.

Table 3 – Nested-PCR mix

Volume per sample	Reagent Name
5 µl	dNTPs (1 mM)*
2,5 µl	Buffer (10x)[a]
1,4 µl	Magnesium chloride ($MgCl_2$) (50 mM)[a]
0,37 µl	Bio-x-act short polymerase (4 U/ml)[a]
2 µl	Primers pair (25 pmol/µl)[b]
3, 73 µl / 11, 73 µl	Water DNase, RNase, Protease-free[c]

* Applied Biosystems [a] Bioline [b] Invitrogen [c] 5Prime

The amplification profile of the nested-PCR reaction comprehended a first step of denaturation and annealing, followed by 35 cycles of extension, denaturing and annealing, and a final extension step, according to the conditions described in Table 4.

Table 4 – *C. trachomatis ompA* Nested-PCR amplification profile

1st PCR (Temperature)	2nd PCR (Temperature)	Time	Cycles
95ºC	95ºC	5 minutes	1x
55ºC	50ºC	1 minutes	1x
72ºC	72ºC	1'25 seconds	
95ºC	95ºC	30 seconds	35x
55ºC	50ºC	30 seconds	
72ºC	72ºC	10 minutes	1x

After the nested amplification, we performed amplicon detection through gel electrophoresis in TAE buffer [1% (w/v) agarose in TAE buffer, SYBR® safe DNA gel stain (10.000x in DMSO; Invitrogen)], for 1 hour at 90 V.

According to the *ompA* nested-PCR results, positive samples were further sequenced. When negatives were from urine samples no further action was done. If the sample was an exudate nested-PCR was repeated and if they remained negative they were excluded.

For sequencing, PCR products were purified using ExoSAP® (Affymetrix) according to the manufacturer's instructions. Briefly, they were subjected to an enzymatic PCR cleanup that consists in two cycles of 15 minutes each (one at 37ºC followed by another at 80ºC) the first with the aim of activating the enzymes responsible for removing contaminants (i.e. excess of primers and dNTPs or single strand DNA) and the second for inactivating enzymes used in the previous step.

After purification, the reaction mix for the sequencing reaction was performed as described in Table 5, using the set of primers described in Annex 3, for a final volume of 9,5 µl, to which 0,5 µl of purified PCR product was added. We used Big Dye technology that is a fluorescent molecule attached to the ddNTPs allowing each one to carry a different color of dye (110).

Table 5 – Sequencing mix

Volume per sample	Reagent Name
2,5 µl	Big Dye *
0,75 µl	Sequencing buffer (5x) *
1 µl	Each primer (5 pmol/µl) [a]
5,25 µl	Water DNase, RNase, Protease-free [b]

* Applied Biosystems [a] Invitrogen [b] 5Prime

The amplification profile for the sequencing reaction is described in Table 6.

Table 6 – Sequencing profile

Temperature	Time	Cycles
96ºC	30 seconds	1x
96ºC	10 seconds	25x
50ºC	5 seconds	
60ºC	4 minutes	

Sanger sequencing was performed by the Unit of Technology and Information (UTI) of INSA. Sequences were further evaluated by us using Chromas Lite program (Technelysium Pty Ltd). For each *ompA* sequence of each strain, we generated a nucleotide alignment with *ompA* sequences from *C. trachomatis* prototype strains representing the various *ompA* genotypes (and variants previously identified in the laboratory). This was done by applying the MegAlign software (DNASTAR) that, besides the alignment, also performs phylogenetic analyses and pairwise comparisons. When new variants were determined, the whole *ompA* typing procedure (nested-PCR plus sequencing) was repeated to confirm the observed mutations.

Statistical analysis was performed in order to determine the odds ratio (a statistical measure of association) between factors; the hypothesis was rejected if the P value for Fisher's exact test was less than 0.05 (111). This analysis was done using GraphPad Prism 6 software (https://www.graphpad.com/scientific-software/prism/).

3. Strategies for quick subtyping LGV strains

The Portuguese legislation regarding the obligatory notification of transmissible diseases (Despacho nº 5681 – A/2014 of 29 April 2014, plus Desclaração de Rectificação nº 609 – A/2014 16 June 2014, plus Despacho nº 15385-A/2016 21 December 2016), establishes that the identification of serovars L1, L2 and L3 is necessary to confirm a case of lymphogranuloma venereum. This can be done by using the routine *ompA* genotyping methodology described in the previously section but this technique takes about a week before getting to a result, and for this reason, whenever a clinical suspicion exists together with the detection of *C. trachomatis*, a quicker methodology should be applied. However, such a test is not yet commercially available. Thus, we intended to implement a methodology that would quickly identify *C. trachomatis* L1, L2 and L3 genotypes. For this, we developed a PCR targeting the *ompA* gene, using sets of primers for amplify a potential region that would differentiate between L serovars. We also implemented a real time-PCR with *ompA* and *pmpH* as target genes, freely available from the web (112), where the *pmpH* gene is used as it clearly separates LGV strains when a phylogenetic tree is constructed based on this gene (31). For both approaches, we used DNA samples from prototype chlamydial strains and from clinical specimens previously *ompA*-genotyped.

3.1 Subtyping LGV: Using sets of primers discriminatory of L1, L2 and L3

Oligonucleotides were designed by the Bioinformatics Unit of the Department of Infectious Diseases of INSA (Annex 4). The PCR mix was prepared as described in the previously section (Table 3) and the PCR profile was identical to the described (Table 4) for the first round of amplification.

After the amplification, we performed amplicon detection through gel electrophoresis in TAE buffer [1,5% (w/v) agarose in TAE buffer, SYBR® safe DNA gel stain (10.000x in DMSO; Invitrogen)], for 30 minutes at 90 V.

To assure the quality of amplicons, PCR products were further purified and sequenced; sequences were analyzed as mentioned in section 2.

3.2 Subtyping LGV: real time-PCR for *C. trachomatis ompA* and *pmpH*

Subtyping of LGV through a real time-PCR approach, was based on a protocol available from the web (112) that implicates *C. trachomatis ompA* and *pmpH* genes. This procedure was done using the oligonucleotides described by authors (2) (Annex 5), and the detection was done by using the SYBR Green technology (LightCycler 480® Sybr green Master Mix I, Roche) according to the manufacturer's instructions. Each set of primers was tested in a COBAS® z 480 Analyzer (Roche). The master mix for each real time-PCR was performed for a total volume of 20 µl as shown in Table 7, to which 5µl of DNA sample was added. The amplification profile is described in Table 8.

Table 7 – Master mix for LGV subtyping real time – PCR

Volume per sample	Reagent Name
12,5 µl	SYBR green*
4 µl	Primer pair (5 pmol/µl)[a]
3,5 µl	LC480 water*

*LightCycler 480® Sybr green Master Mix I (Roche) [a] Eurofins Genomics

Table 8 – Amplification profile for LGV subtyping real time – PCR

Temperature	Time	Cycles
Pre-incubation		
95ºC	10 minutes	1x
PCR		
95ºC	15 seconds	40x
60ºC	1 minute	
Melting		
95ºC	15 seconds	1x
60ºC	20 seconds	1x
95ºC	Continuous	1x
Cooling		
40ºC	30 seconds	1x

4. Homopolymeric tracts counts and bioinformatics analyses

4.1 Study design: selection of homopolymeric tracts and *C. trachomatis* positive DNA samples

For the study of potential targets of phase variation in *C. trachomatis*, we were supported by the Bioinformatics Unit of the Department of Infectious Diseases of INSA, which has been developing a strategy focused on sequencing, through a high throughput next-generation sequencing (NGS) technology with amplification products (amplicons) obtained with PCR schemes targeting homopolymeric tracts. Using whole-genome sequences available on Genbank, we firstly looked for regions in the *C. trachomatis* genome displaying homopolymeric tracts [poly (Ns)] with base counts equal or above to eight. Additionally, whole-genome sequence data from the host laboratory previously studied *C. trachomatis* population (104) was systematically analyzed for the presence of intra-strain heterogeneous and non-heterogeneous homopolymeric tracts. Altogether, this allowed us to select 20 genome-dispersed homopolymeric tracts identified to be likely prone to allelic variation (targets) and 4 homopolymeric tracts highly conserved within *C. trachomatis* intra-populations (potential future controls for each base type). The poly (N) selection have also taken into account the poly (N) location in relation to the potentially affected gene (upstream or coding region) and also on the function of the gene (Annex 6).

Although it is our goal to perform a future large-scale study involving multiple DNA samples obtained from *C. trachomatis* positive clinical specimens associated with distinct anatomical sites (genital versus anorectal mucosa) and distinct *C. trachomatis* *ompA*-genotypes, in the present study, we performed a preliminary assay of validation/optimization of the novel approach by evaluating the selected 24 poly (N)-targeting amplicons for 10 clinical samples (Annex 7). In particular, we specifically aimed at: i) implementing PCR schemes targeting each one of the 24 selected poly (N) s; ii) comparing the results obtained through the amplicon-based approach with previous

results generated through WGS; iii) evaluating the impact of extending the PCR run on the poly (N) counts profile; iv) getting insight on both the profile of intra-patient variation of putative phase-variable loci and the existence of the tropism-specific homopolymeric lengths.

4.2 Amplicon-based NGS and bioinformatics analyses

PCR primers targeting each one of the 24 selected poly (Ns) were designed using Primer Select of DNASTAR (Lasergene 9 Core Suite) (see Annex 8) and oligonucleotides were ordered complemented with Illumina adapter sequences at their 5' ends (see Annex 9).

A PCR technique was used to amplify each homopolymeric region, using the DNA polymerase KAPA HiFi HotStart ReadyMix PCR Kit from KAPA Biosystems, an enzyme recognized by minimizing polymerase induced errors, according to the manufactures' instructions. We prepared PCR mixes (Table 9) for a total volume of 22,5 µl to which 2,5 µl of each DNA sample were added. Each PCR promotes an amplicon of about 250 bp.

Table 9 – PCR mix for homopolymeric targets

Volume per sample	Reagent Name
12,5 µl	KAPA HiFi Polymerase*
2 µl	Primers pair (5 pmol/µl)[a]
8 µl	Water DNase, RNase, Protease-free[b]

* KAPA Biosystems [a] Invitrogen [b] 5Prime

The amplification profile comprehended a first step of denaturation, followed by 35 or 45 cycles of denaturing, annealing and extension, ending with a final extension step, according to the conditions described in Table 10. Annealing temperatures were optimized for each PCR scheme.

Table 10– Amplification profile for homopolymeric regions

Temperature	Time	Cycle
95ºC	3 minutes	1x
95ºC	30 seconds	35 or 45x
Annealing temperature [a]	30 seconds	
72ºC	30 seconds	
72ºC	5 minutes	1x

[a] This temperature was adjusted for each PCR scheme

Amplicons generated for all gene targets for each clinical isolate were pooled in a single tube and sent to the UTI – INSA, where they were further cleaned-up, independently indexed and subjected to cluster generation and paired-end sequencing (2 x 150 bp) in an Illumina MiSeq equipment (Illumina Inc., San Diego, CA, USA), according to the manufacturer's instructions. In order to evaluate the variation within DNA homopolymeric tracts, the Bioinformatics Unit of the Department of Infectious Diseases of INSA applied an in-house Python script [described in (113)] that enables the extraction and counting (directly from raw reads, both forward and reverse) of DNA sequences that are contiguously flanked by two conserved, user-defined, small DNA strings, as represented in Figure 5 and previously described by the host laboratory (113). As such, after defining conserved regions flanking each one of the 25 homopolymeric tracts under study, we have applied the script to determine the precise relative frequency of clones carrying specific base counts within a given sample.

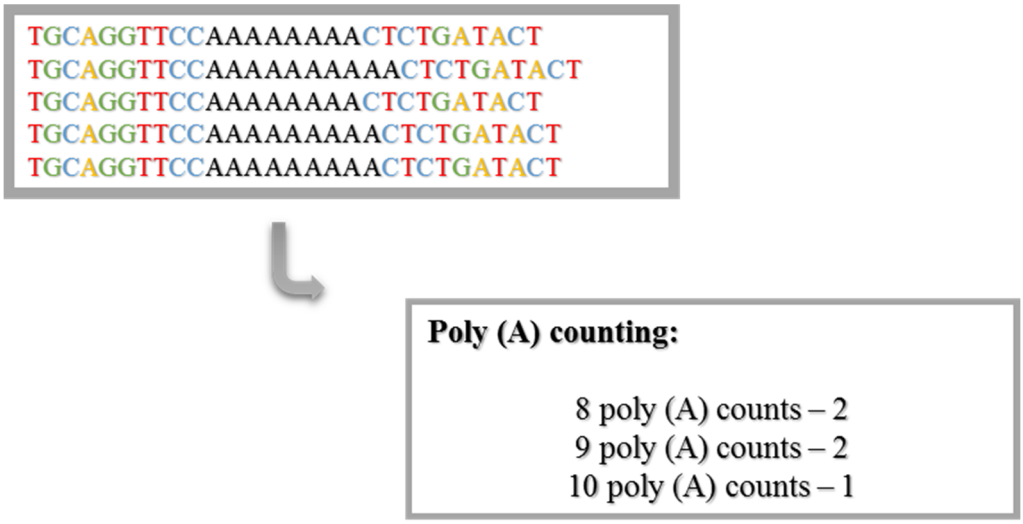

Figure 5 – Schematic representation of homopolymeric traits counts

III. Results and Discussion

1. Cell culture and *C. trachomatis* typing

1.1 Cell culture, the old gold standard

As previously mentioned, the cell culture used to be the gold standard technique for the diagnosis of *C. trachomatis* (81). However the requirements of the cell culture technique and for maintaining the bacteria alive until inoculation, led to the implementation of new techniques more affordable for diagnosis. These facts become obvious from our results; in fact for a total of 29 strains submitted to propagation in cell culture (7 ATCC prototype strains and 22 clinical isolates), only 4/7 (57%) ATCC and 6/22 (27%) clinical isolates were able to grow (Figure 6A and Figure 6B, respectively).

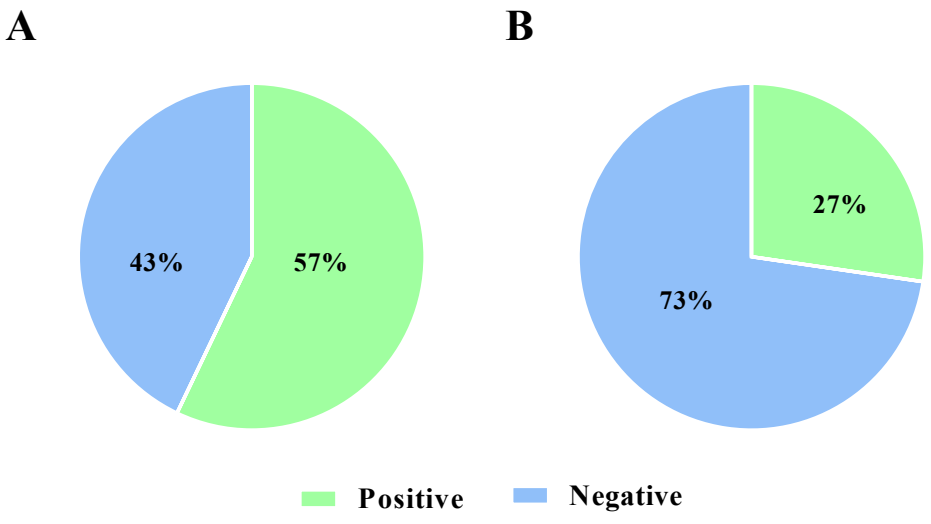

Figure 6 – Results from propagating *C. trachomatis* strains in cell culture. Panel A represents ATCC prototypes strains and panel B characterizes clinical samples.

The ATCC strains were cultured with the aim of maintenance of prototype strains in the laboratory, providing an opportunity for learning cell culture and chlamydial propagation. ATCC samples were expected to have high bacterial loads; however, from our experience, almost half provided negative cultures. This could be related with deficient storage conditions, due to deficient freezing (above -80ºC) or broad temperature variations of the ultra-freezer; thus, the storage (even for short periods of time) under those non-optimal conditions might have contributed to the decrease of the infectivity. Clinical strains were cultured as a contribution for the Portuguese collection of *C. trachomatis*. The success of chlamydial culture relies on the accomplishment of several factors that can easily fail, from which transport in appropriate media, temperature and storage under -80ºC, stand out. Furthermore, samples that are selected for culture have prior been considered positive through commercial nucleic acids amplification techniques, which are recognizably more sensitive. Even the nested-PCR technique used for *ompA*-genotyping provides more positive results than culture, as 12/16 (75%) specimens that were negative in culture revealed positive upon nested-PCR. The four negative specimens, both in culture and *ompA*-nested-PCR, cannot be fully excluded from being true negatives; however, the lack of sensitivity of these methodologies (86,114,115) is the most probable explanation.

Live *C. trachomatis* prototype and clinical strains are considered of great interest for research studies, and the prior have been long used for such purposes; however, recent (104) whole genome sequencing studies upon which the genome of clinical strains was sequenced before and after several cycles of propagation in cell culture, revealed that *C. trachomatis* genome rapidly evidences several changings. This phenomenon has been considered (104) as an adaptation to *in vitro* life conditions, whereby some genes can be lost because they would no longer be needed (i.e. to provide nutrients that are made available through culture media), or because of their involvement in the *in vivo* infectious process and would no longer be required when the microorganism has no host immune system to fight against. Despite these considerations, much of chlamydial research will still depend on live strains and for this reason, the collection of clinical strains that are infecting humans' remains obligatory to a national reference laboratory.

1.2 Characterization of *C. trachomatis* clinical strains

We got a collection of 278 *C. trachomatis* positive samples to perform *ompA* – genotyping, and we were successful for 245 (88,1%) of them (Figure 7).

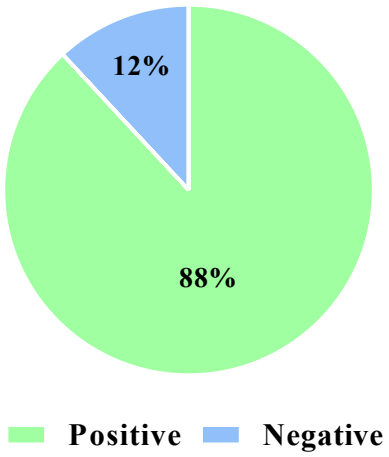

Figure 7 – Rate of *C. trachomatis ompA*-genotyped strains

From the 278 *C. trachomatis* strains, 2 were obtained from the conjunctiva of newborns, due to mother to child transmission during delivery upon passage through an infected birth canal. These 2 samples will not be considered for analysis regarding to gender and age. Detected from women were 112 (41%) specimens, whereas 164 (59%) were detected from men samples (Figure 8).

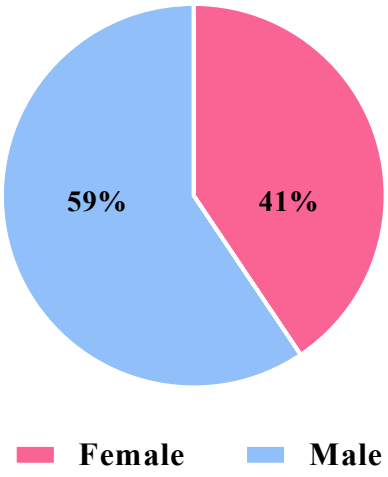

Figure 8 – *C. trachomatis* strains by gender

Unlike most chlamydial studies, that comprehend a majority of women, a gender considered more prone to search for medical care, namely gynecological routine examination, our population involved more men. This relates to the attendees of the STI clinic from which most of the samples are taken for *C. trachomatis* diagnosis at INSA, mostly constituted by MSM, obviously leading to a bigger frequency of the male gender. In addition, laboratories that are currently sending positive chlamydial specimens for *ompA*-typing at INSA, also provide *C. trachomatis* screening to MSM, contributing to a collection where strains detected in men are more represented.

We evaluated 276 strains according to two age groups, one including strains from people with less than 25 years old and the other including people having 25 years old or more. People with 25 years old or more provide 173 (63%) strains and people with less than 25 years old, 103 (37%) strains (Figure 9).

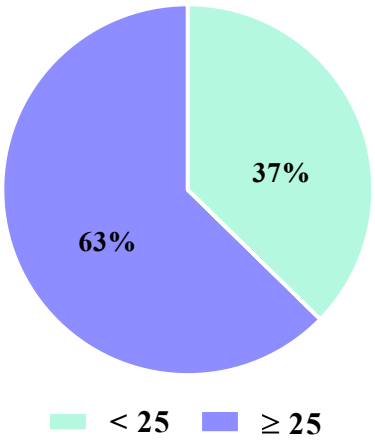

Figure 9 – *C. trachomatis* strains by age

Strains were *ompA*-genotyped from infections detected in persons from 14 to 61 years old, median 27 years old and media 29 years old.

Most of our strains were detected in persons aged 25 years old or more, most probably due to the age of the attendees of STI clinics to which INSA provides routine *C. trachomatis* diagnosis. The fact that *C. trachomatis* infects people of older ages shows to the evidence that, despite the development of some immunity through several episodes of infection (116), it does not provide enough protection to avoid the occurrence of new episodes of chlamydial infections. However, in terms of public health, the infection is more important for people with less than 25 years old due to the clinical sequelae related to infertility.

Figure 10, represents the distribution of positive *C. trachomatis* samples per anatomical site. Urine was the most common biological sample providing positives, with a total of 124/278 (44,6%), followed by cervical swabs with 76/278 (27,3%), and anorectal swabs with 56/278 (20,1%). Less usual were urethral swabs (nowadays replaced by urine, particularly in men, due to its non-invasive character) with 13/278 (4,7%) samples, and two sorts of specimens where chlamydial diagnosis is rarely asked at INSA, the oropharynx, 7/278 (2,5%) and the conjunctiva 2/278 (0,7%).

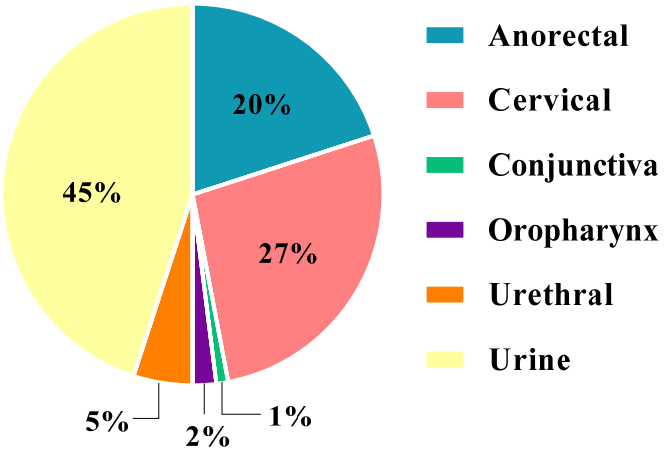

Figure 10 – Frequency of biological samples positive for *C. trachomatis*

Although urines became the sample of choice, in particular for men suspected of urethritis, for women it has not replaced the cervical swab, as this provides more sensitivity even when using the modern commercial nucleic acid amplification tests (6). The third most common sample that provided positive tests were the anorectal swabs that are collected mainly in MSM, for detecting chlamydial proctitis. The transmission of *C. trachomatis* from an infected mother to the newborn during delivery has been established in 50-70% (6), and newborns with conjunctivitis should be tested for *C. trachomatis*. However, that type of diagnosis is usually performed at the hospital, rarely at INSA; thus, the few samples existing in this collection came from a pediatric hospital in Lisbon. In the adult, chlamydial conjunctivitis can also occur (54), but the laboratory diagnosis of *C. trachomatis* in the adult conjunctiva is barely done at INSA. Similarly to the conjunctiva, also oropharyngeal chlamydial infections are probably underdiagnosed, as very few diagnoses were asked for that anatomical location. Nonetheless, oropharynx swabs should be routinely tested whenever the individual has unprotected oral intercourse (117).

Sanger sequencing of *ompA* amplification products allowed the identification of 12 genotypes among the 245 positive samples. The most represented genotypes were E (n=55; 22,5%), D (n=47; 19,2%), F (n=40; 16,3%) and G (n=38; 15,5%). Less represented were genotypes L2 (n=26; 10,6%), J (n=15; 6,1%), I (n=12; 4,9%), K (n=7;

2,9%) and H (n=3; 1,2%). For two samples (0,8%) the exact *ompA* genotype could not be identified, as chromatograms exhibited several nucleotides overlapping, thus they were considered as mixed infection (Figure 11).

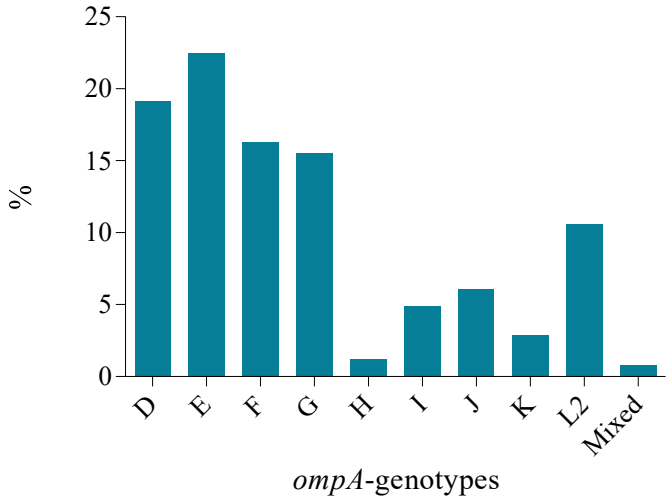

Figure 11 – Frequency of the different *ompA*-genotypes among the studied *C. trachomatis* strains. D includes (33/47) 70% Da; all I were from Ia variant; J includes (2/15) 13% Ja; L2 includes (6/26) 23% L2b.

Along the years, *ompA*-genotype frequency has changed, D, E and F remain the most frequent *ompA*-genotypes, as described in previous work of the host lab (27,118,119), and a common finding in studies from other countries (120–122). In fact, E and F have been considered the most successful *C. trachomatis* genotypes (123). However, there have been some changes in *ompA*-genotypes frequency at INSA. The typing methodology has changed from the Restriction Fragment Length Polymorphism (RFLP) to *ompA* sequencing; however, methodology should not be considered the reason for *ompA*-genotype frequency changing, as RFLP provided results that were later confirmed by sequencing at INSA (data not shown). For instance, the high frequency of serovar H by the end of 1990s (118) was related with transmission within a sexual network; so, when it stopped existing, the frequency of H decreased (27,119). On the contrary, genotype G became the fourth most frequent since the beginning of 2000 (27,119), probably because many of the men that tested positive for *C. trachomatis* were MSM and genotype G has been associated to this population (76). Finally, L2 genotypes that were rarely found until 2004, when LGV reemerged in developed countries

(27,118,119), became quite common, most probably due to the high number of MSM attending to STI clinics from which samples are tested for *C. trachomatis* at INSA; in fact, most of the LGV cases described all over the world are described in MSM and are caused by genotype L2, while L1 and L3 cases remain rare (65).

ompA-genotyping is not discriminative enough to perform contact tracing; however, among the *C. trachomatis* strains genotyped through our study, 12 were identified in heterosexual individuals who were sexual partners, and 3 couples were indeed infected by the same *ompA*-genotype. For the remaining 3 couples only one partner was infected (Ja) while the other revealed negative for *C. trachomatis*; in another couple, one was infected by genotype D while the partner was infected by F; finally one person was having two partners simultaneously, and in this case two were infected by genotype K while one was infected by genotype J. Despite the lack of contact tracing power, *C. trachomatis* can easily be transmitted through sexual intercourse; thus, partner notification constitutes a useful prevention tool, contributing for early screening and treatment and therefore for breaking transmission chains.

For the exposed, molecular epidemiology studies involving *C. trachomatis* typing are of great importance to provide knowledge and awareness to health community, namely to evaluate the reasons for changes in genotype frequency that might be related with specific clinical features and/or population.

ompA sequencing allowed the detection of genotype variants previously described, as well as new variants. A variant corresponds to nucleotide changings in relation to the described for the same position in the sequence of the prototype strain of a given genotype (27). It should be noted that prototype strains (also designated as reference strains) did not get that status because they are the most common circulating type; reference strains became prototype strains because they were the first described for each *C. trachomatis* type and became ‘popular’ among researchers because the same strain could be used for several research approach and applications and results would be comparable. So, it might happen that a reference strain is itself, a quite rare or no longer circulating variant. In fact, in our study, the most common circulating *ompA*-genotype corresponds to a variant; it is the case of clinical strains H, and more obvious of D (where

Da stands out as the most frequent variant), and G (where mutation at nucleotide 1003 was the most frequent variant) (Figure 12).

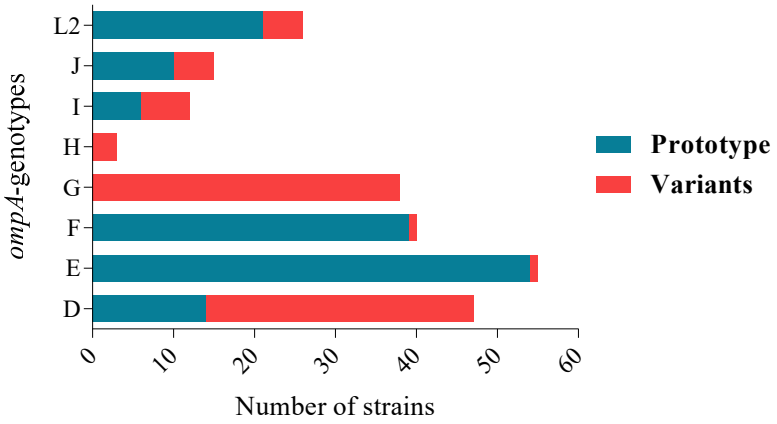

Figure 12 – Number of *ompA* variants among detected *C. trachomatis* genotypes

A total of 6 new variants were detected in genotypes D (all from the Da variant), F, G and L2 (one from the prototype and two from the L2b variant). Figure 12 exhibits the number of variants identified through this study, and as stated above, all strains of genotypes G and H appear as variants. We could not find any particular reason for this finding, but we could speculate that these genotypes still did not get the best fitting for their major antigen, MOMP, and thus, they keep undergoing an adaptation process, in contrast to genotypes E and F among which the number of variation is much lower. It is of note that 40% (10/23) of the mutations involve epitopes for B (n=9) or T cells (n=1), suggesting implications related to evasion from the host immune system (Table 11). The adaptation that any microorganism must undergo, as a consequence of any mutation, has a cost in terms of energy consumption. Thus, the maintenance of a mutation only occurs when it beneficiates the microorganism, namely by allowing it to escape from defense mechanism of the host (27); this is the putative reason why some mutation persist while others are soon eliminated from a microbial population.

Table 11 describes all the variable sites identified in the present study, either new (this study) or previously described.

Table 11 - Nucleotide sequence variation within *ompA* in relation with the respective prototype strain [Table based on (27)]

Genotype (no. Of variants)[a]	No. of variable sites[b]	MOMP region[c]	Nucleotide position[d]	Amino acid change[e]	Type of change[f]	No. of isolates[g]
D (33)	5	CDII	436	Leu → Phe		2/33
		CDIII	636	Asn → Ser[1]		20/33
		VDIV	939	Asn → Ser[1]	B-cell	20/33
		VDIV	976	Ala → Thr[1]	B-cell	1/33
		VDIV	997	Thr → Ala[1]	B-cell	20/33
E (1)	1	VDII	514	Ser → Ala[1]	B-cell	1/1
F (1)	2	VDIII	773	Ala → Val		1/1
		VDIV	1003	Thr → Ala	B-cell	1/1
G (38)	7	CDI	228	Asp → Asn[1]		1/38
		CDIII	700	Glu → Gln[1]		1/38
		VDII	487	Gly → Ser[1]		4/38
		CDIV	921	Gln→ Gln		1/38
		VDIV	985	Ser → Gly[2]		1/38
		VDIV	1003	Ser → Ala[1]		32/38
		VDIV	1003	Ser → Thr[1]		2/38
H (3)	2	VDI	272	Asn → Ser[1]	B-cell	3/3
		CDIV	850	Ala → Val[1]	T-cell	3/3
I (6)	1	VDII	526	Ile → Val[1]		6/6
J (5)	1	CDII	369	Ala → Thr[1]		5/5
L2 (7)	4	VDII	485	Asn → Ser[1]	B-cell	6/7
		VDII	494	His → Arg	B-cell	1/7

	VDII	515	Lys → Thr	B-cell	1/7
	VDII	517	Leu → Ile		1/7

[a] **Genotype (number of variants)** – Genotypes with the number of variants found in this work;

[b] **Number of variable sites** – Number of variable sites found in this study that had been previously identified or new;

[c] **MOMP region** – Indicates the region where variable sites are located. CD – Conserved region, VD – Variable region; for both I, II, III or IV;

[d] **Nucleotide position** – based on the alignment of the strains from the same genotype;

[e] **Amino acid change** – comparison of the clinical variant in relation to the prototype for each genotype. Ala – Alanine, Arg – Arginine, Asn – Asparagine, Asp – Aspartic acid, Gln – Glutamine, Glu – Glutamic acid, Gly – Glycine, His – Histidine, Ile – Isoleucine, Leu – Leucine, Lys – Lysine, Phe – Phenylalanine, Ser – Serine, Thr – Threonine, Val – Valine. [1] Described in *Nunes et al*, 2009 (27) [2] Described by *Li, C. L. et al*. The molecular epidemiology of genital *Chlamydia trachomatis* in China. Unpublished. 2005;

[f] **Type of change** – Mutations that occurred within B-cell epitopes (B-cell); mutations that occurred within T-cell epitopes (T-cell);

[g] **Number of isolates** – The number of isolates sharing each variable site.

Among the 23 amino acid changes, one led to a synonymous mutation in a conserved domain, while the remaining 22 involved non-synonymous mutations (Table 11). These phenomena reflect the plasticity of MOMP in response to environmental pressure. Despite being hard to define impact of these changes on a whole protein level, given it complexity (27), aminoacid changing may have implications in the conformation of MOMP; in fact 16/22 (72,7%) of the non-synonymous mutations occurred in variable domains of the major chlamydial antigen, and may induce changes in specific epitopes, which could contribute for evading from host immune defense mechanisms.

We also wanted to evaluate *ompA*-genotype regarding the anatomical sites of infection. We defined three groups: genital group included all samples obtained from the cervix, urethra (including urines) and conjunctiva, assuming the acquisition of the infection through vaginal intercourse or transmission to newborn during labor (due to contact between the eye and the cervix during passage by the birth canal, as already mentioned); anorectal group included anorectal samples, all from men, all related to anal intercourse in MSM; and oropharynx group, the smallest, including oropharynx samples related to oral intercourse. Figure 13 shows that D-K were more often found in chlamydial

infections acquired through vaginal intercourse, while LGV genotypes were more frequently detected, 97-fold higher (OR, 95% CI 21,6-438,7) in chlamydial infections acquired through anal intercourse in MSM; a population among which, G was the most common *ompA*-genotype in oropharynx samples. Genotype G has been associated with chlamydial infections among MSM but the reasons underlying this finding are yet to be fully understood (76).

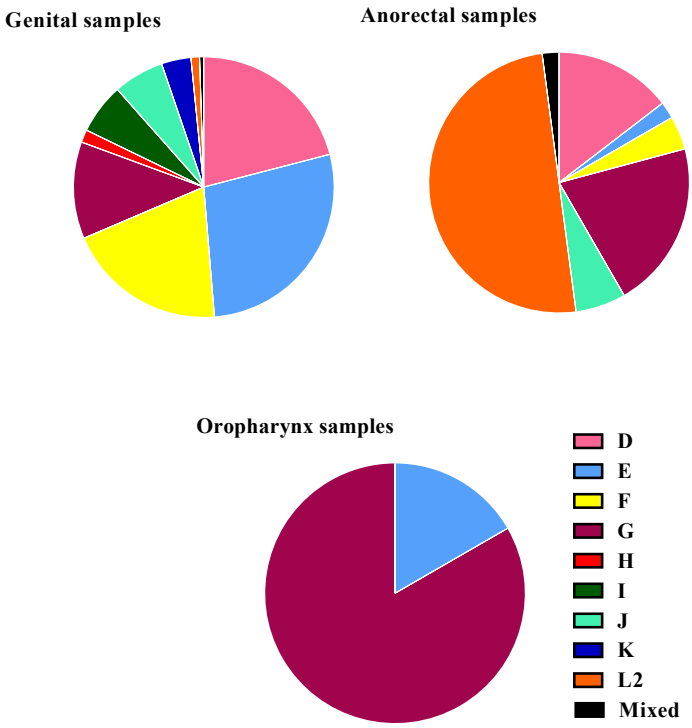

Figure 13 – *C. trachomatis ompA*-genotype distribution according to anatomical site of infection

We tried to understand if there was a relationship between age and the most common *ompA*-genotypes. In fact, Figure 14 reveals that L2 strains are more frequent in people aged 25 years old or more; the reasons underlying this observation could not be disclosed through the present study, but could better be related to social features of the MSM population, than to biological characteristics of *C. trachomatis* L2 genotypes or of infected individuals, being 9-fold higher to detected a LGV positive case in individuals aged more than 30 years (OR, 95% CI 3,3-25,4).

For D-K genotypes, the distribution is relatively similar between age groups for most of them, with the exception of genotypes H and K that were only detected in

individuals having 25 years or more; however, the number of strains for each one of them, 2 and 7 respectively, turns hard to value this observation.

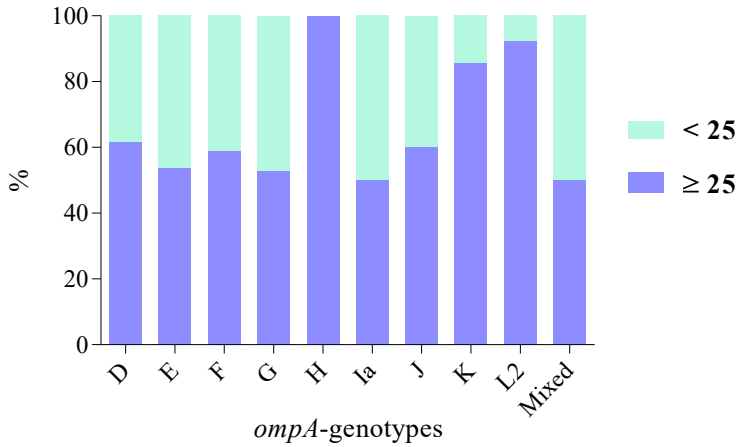

Figure 14 – *C. trachomatis ompA*-genotype frequency per age group

2. Strategies for quick subtyping LGV strains

In 2014 *C. trachomatis* was included in the Portuguese list of transmissible diseases under obligatory notification by law (Despacho nº 5681 – A/2014 of 29 April 2014 plus Declaração de Rectificação nº 609 – A/2014 16 June 2014, recently updated by Despacho nº 15385-A/2016 21 December 2016). This list further specifies the need for identifying *ompA*-genotypes L1, L2 and L3 to confirm any LGV case but, at present, there is no commercial test available for providing that. To our knowledge, there is a single commercial test able to specifically detect and differentiate *C. trachomatis* LGV from *C. trachomatis* non-LGV types, CLART® STIs (Genomica). However, this distinction is not based on the *ompA* gene but instead, on the *pmpH* gene (31,124); as the chlamydial genome is prone to recombination (125) it could happen that a strain might have a *pmpH* gene similar to LGV strains while its *ompA* is A-K. So, this commercial test does not fulfill the requirements defined by the Portuguese law to confirm a LGV case, because it is not based in *ompA*-genotyping. Of course, the *ompA*-genotyping method described and used throughout our study, can answer to that legal requirement; however, it is a laborious and time consuming methodology, based in a nested-PCR followed by sequencing, a process usually takes at least one week, but can also take more than one month if the nested-PCR result is negative (sometimes leading to new nucleic acid extraction procedures, or just the repetition of the nested-PCR reaction) or if the sequencing process does not provide sequences of enough quality to allow *ompA*-genotype determination. Thus, we tried to implement a faster and reliable methodology that would overcome regular *ompA*-genotyping difficulties to quickly and specifically identify L1, L2 and L3 strains.

For the first approach, based on conventional PCR, only one clinical strain provided acceptable results, as for clinical samples the sensitivity provided by the PCR was clearly insufficient, even when using primer-pair “L2”, which provided better results with prototype strains (Figure 15). Moreover, this approach would not differ greatly from the routine *ompA*-genotyping methodology in terms of time consumption as it would still depend on successful sequencing results.

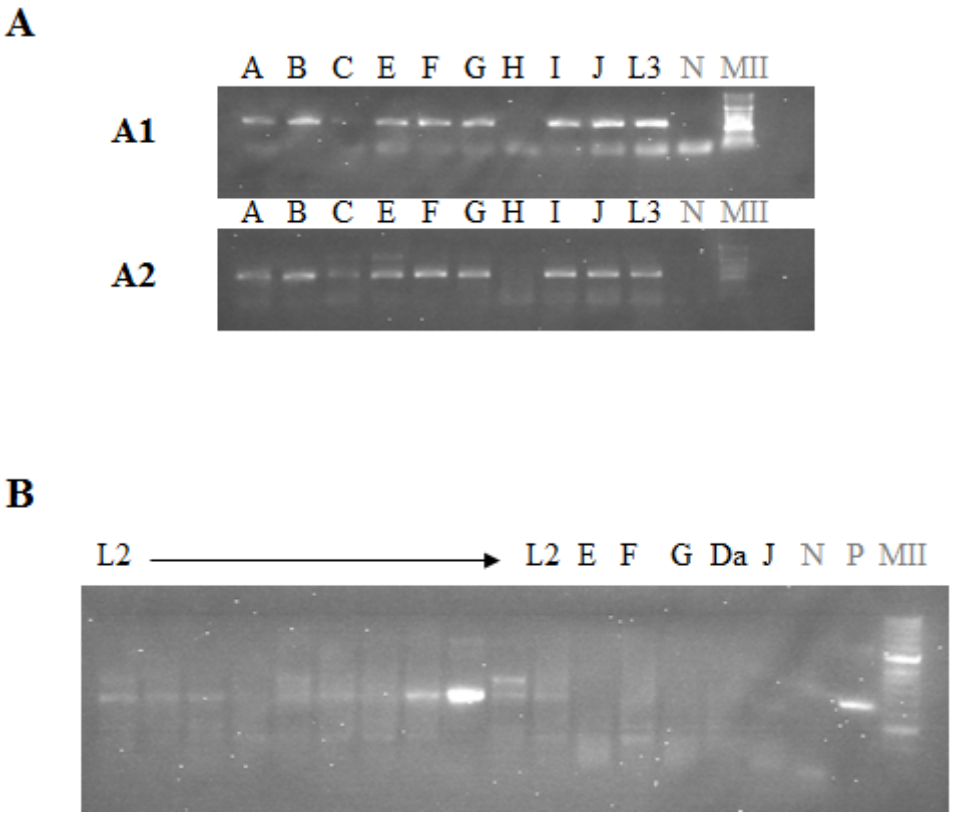

Figure 15 – Conventional PCR approach for detecting L1, L2 and L3. Panel A1 and Panel A2 show the results when using prototype ATCC strains and primer pair "L1" and L2", respectively, for strains A-Har13, B-Har36, C-TW3, E-Bour, F-IcCal3, G-UW57, H-UW4, I-UW12, J-UW36, L3-404. Panel B represents the results when using clinical strains of *ompA*-genotype L2, E, F, G, Da and J with primer pair "L2".

In order to turn around this problem, we also implemented a real time PCR approach that should increase sensitivity and decrease time until result. We used a technique described by others (112) without using the specific probes. This procedure could be considered promising, as amplification curves with a convenient profile at an adequate cycle threshold (Ct) were obtained Figure 16.

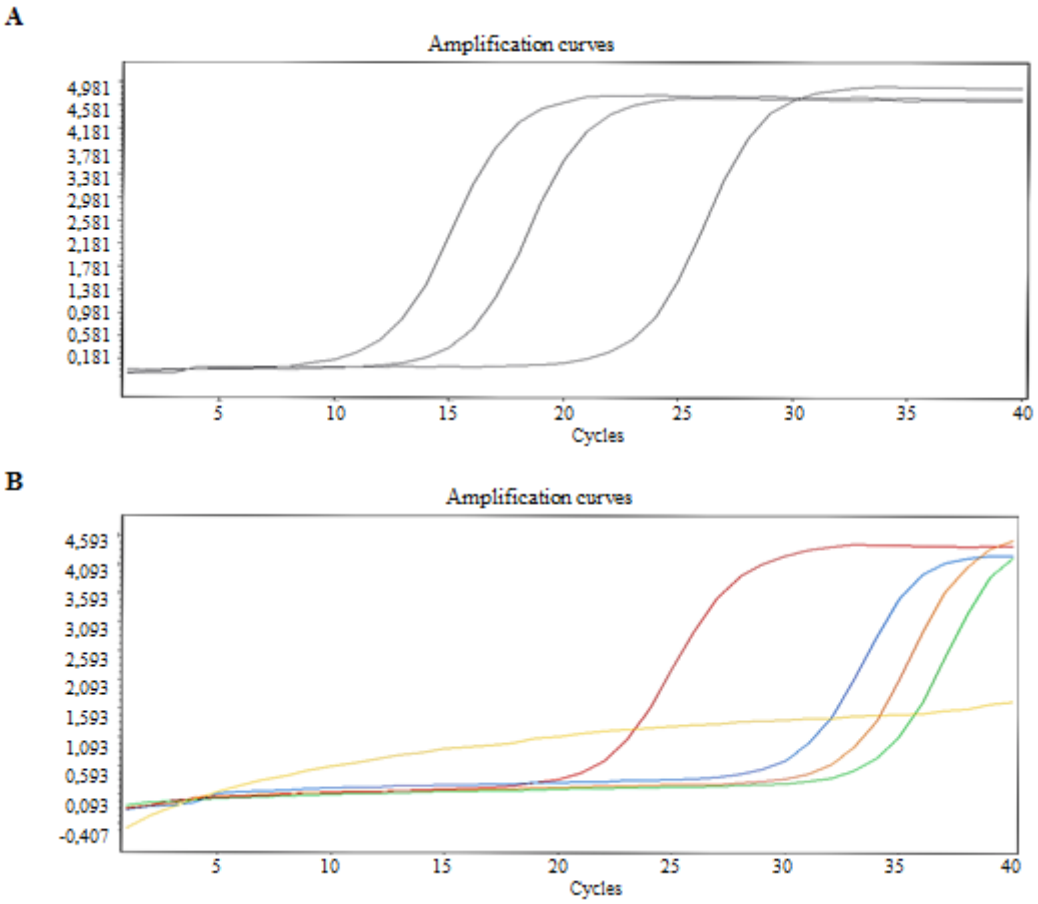

Figure 16 – Results for real time PCR assay. In panel A for L1-440 prototype strain. In panel B for clinical strain of genotype J: detection of *C. trachomatis* (red), LGV (blue), L1, (orange), L2 (yellow) and L3 (green).

Considering that amplification could be observed for all primer sets, this real-time technique could be considered of low specificity and unable to identify LGV strains (as we used a J genotype to test it) and consequently, it could be considered inadequate to distinguish L1, L2 and L3 (Figure 16, panel B). In fact, positive amplification results for several dilutions of different prototype and clinical genotypes were perceived (data not shown); however, the lowest Ct was always observed for the specific set of primers. Thus, we may expect that when the real time is further developed at INSA, using the specific probes, the specificity will improve to proper levels.

The design of a molecular biology technique for specifically distinguish *C. trachomatis ompA*-subtypes constitutes a huge challenge, since the differences among

LGV subtypes are just a few mutations, limiting the choice of primer design to sequence areas where specific mutations for each subtype are located. So, on one hand, there is a loss of sensitivity because primer design itself is far from the ideal conditions for selecting the best primer pair for optimal results, and on the other hand, there is always the risk of unspecific annealing and loss of specificity. The later should be attenuated by increasing the temperature of the annealing step of the PCR profile and by using specific probes that should select only the correct PCR product.

3. Phase variation in *C. trachomatis* – Preliminary study

3.1 Validation of the approach

We aimed to take advantage of the next-generation sequencing (NGS) methodologies to get insight on the genetic heterogeneity affecting homopolymeric tracts potentially driving phase variation in *C. trachomatis*. The study of intra-strain variability *in vivo* is of importance, namely as a means to identify potential phase-variable targets associated with bacterial adaptation in the context of infection. With this purpose, we participated in the development of a strategy focused on a deep sequencing of amplification products (amplicons) obtained with PCR schemes targeting homopolymeric tracts [poly (N)]. In a first stage, we have selected 24 poly (N) study targets by inspecting whole-genome data available at both GenBank and the host laboratory. PCR schemes targeting each selected poly (N) were then designed and optimized (see Material and Methods section), and further applied to DNA samples directly collected from *C. trachomatis*-positive clinical specimens. The resulting amplicons were subjected to high throughput NGS (in an Illumina system). In order to start validating the novel approach, we compared the profiles of poly (N) counts previously obtained at INSA by using WGS, with the ones collected with the amplicon-based technique. In this specific assay, we were able to compare the heterogeneity of 10 distinct poly (N) for two strains (*ompA*-genotypes D and E) (Figure 17). In addition, we

have also evaluated the "amplicon-derived" results when increasing PCR cycles from 35 to 45 (Figure 18). This protocol variation related to the need of potentiating the sensitivity of the assay, and for obtaining a sufficient amount of amplicons copies for sequencing. This is relevant, considering that in a biological specimen the amount of human DNA will be disproportionately higher in comparison to chlamydial DNA. Moreover, although we used a high fidelity polymerase specifically advisable for amplifying repetition regions without sequence errors, in the 45 cycles PCR approach, we exceeded the manufacturer's recommended limit for samples with weak DNA component (35 cycles); so, this comparison also aimed at getting insight on the impact of extending the PCR run on the poly (N) counts profile.

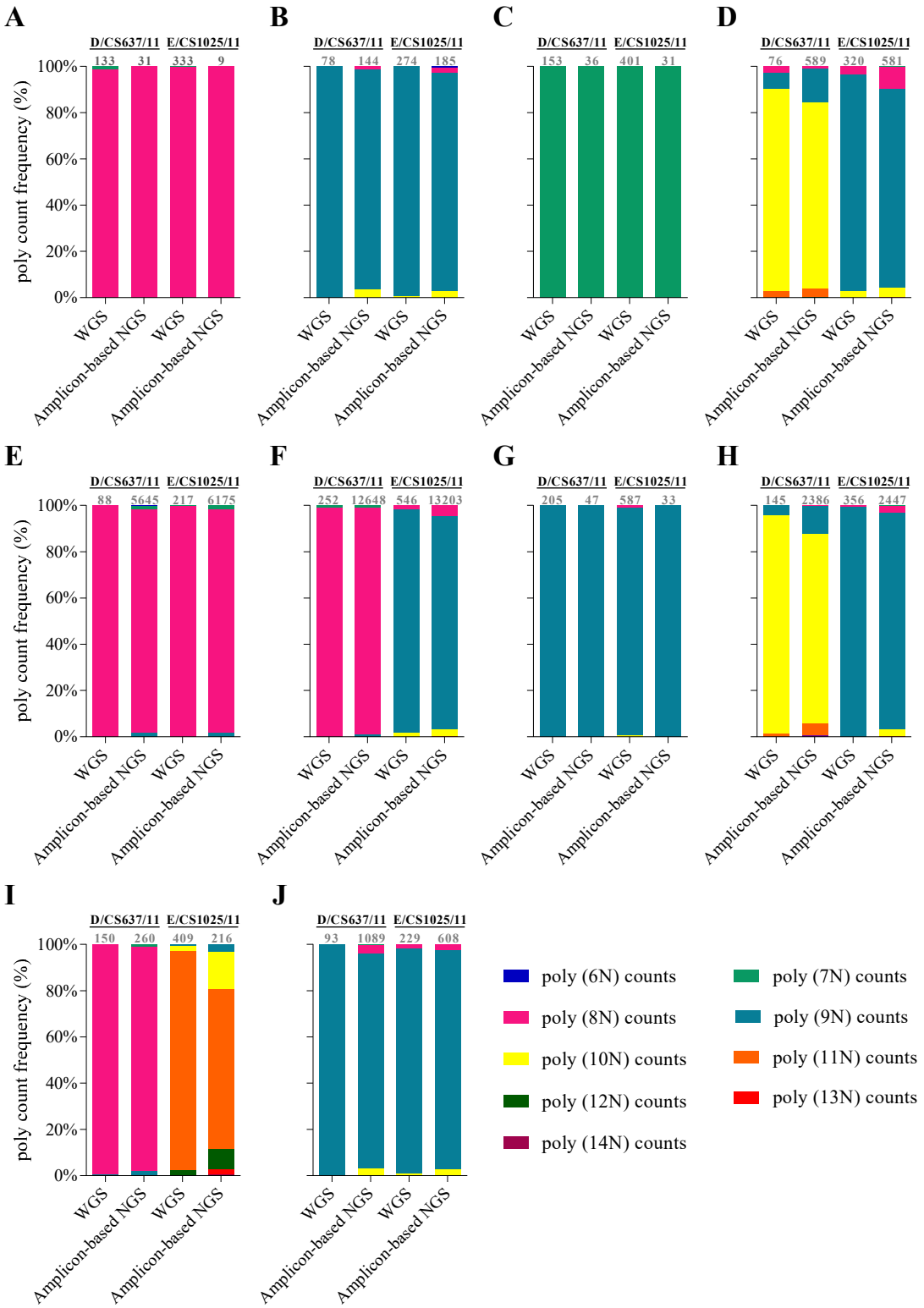

Figure 17 – Comparison between WGS and amplicon-based NGS techniques. For each panel, two clinical strains are compared: *ompA* genotypes D and E. The total number of reads counted for each poly (N) is written on the top of each column (in grey). Panel A shows the results for the poly (N) potentially affecting the gene CT043, while panel B for gene CT054, panel C for gene CT134, panel D for gene CT166,

panel E for gene CT226, in panel F for gene CT326, in panel G for gene CT414, in panel H for gene CT541, in panel I for gene CT561 and in panel J for gene CT694. The results presented for the amplicon-based approach were obtained with 35 cycles of PCR.

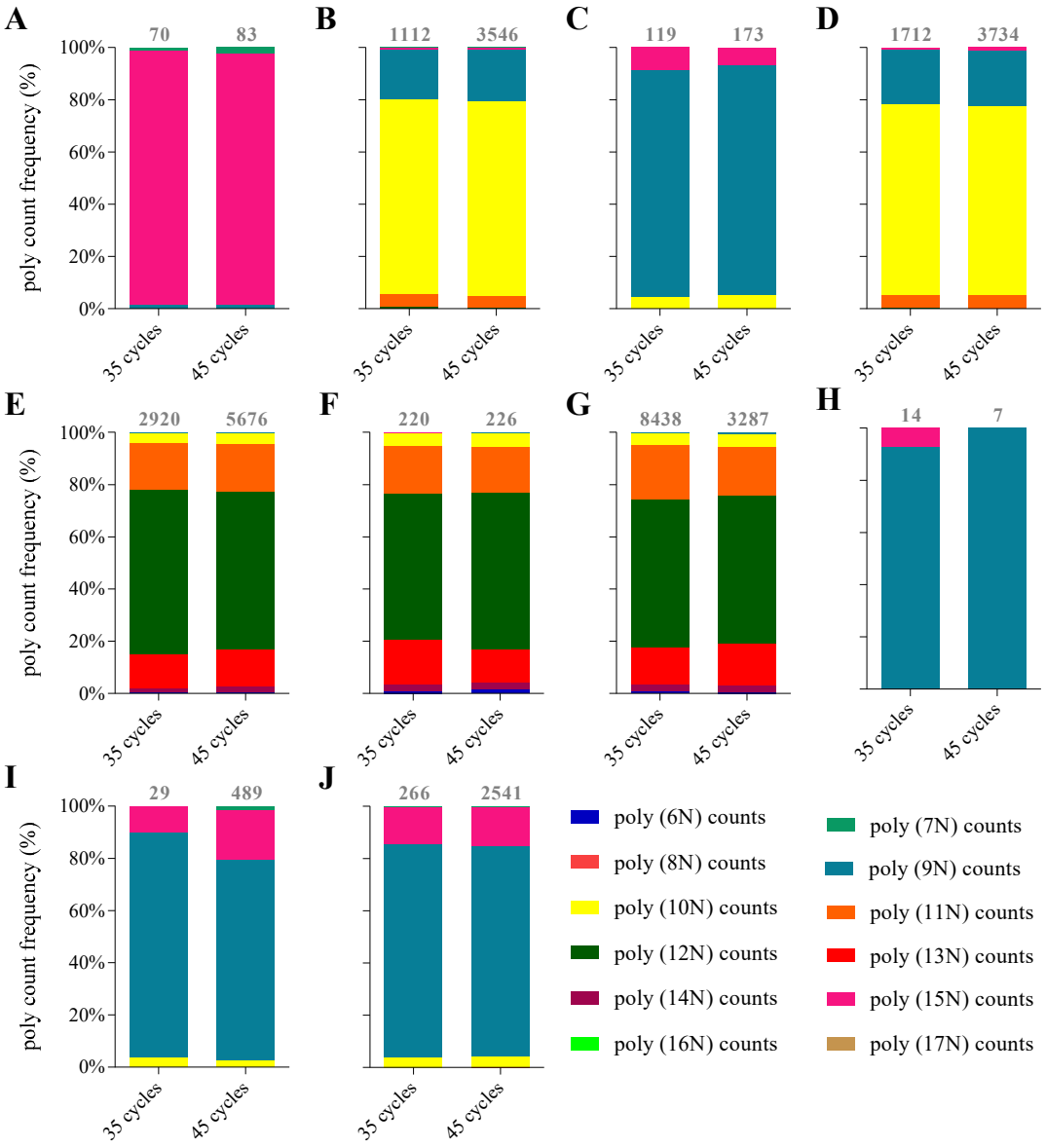

Figure 18 – Comparison between 35 and 45 cycles PCR. Each panel shows the result for 35 cycles and 45 cycles. On top of each graph bar it is stated the total number of reads counted for poly (N) (in grey). Panel A shows the results for gene CT085; panel B, C and D for gene CT166 for different strains; in panel E, F and G for gene CT533 for different strains; in panel H for gene CT590; and panel I and J for gene CT871 for different strains

Taking all these results together, the WGS and the amplicon-based approaches provided very similar results (Figure 17), indicating that the latter, which is a highly customized technique that enables ultra-deep sequencing, seems not impact the relative counts of homopolymeric regions; thus, being appropriate for evaluating phase variation-driven intra-strain genetic heterogeneity. Also, the increase of ten PCR cycles contributed to enhance sensitivity without introducing errors with impact on the counts estimates (Figure 18). Importantly, the most abundant poly (N) count obtained for a given strain was always the same, regardless of the approach employed (WGS or PCR with 35/45 cycles). Moreover, slight differences were essentially observed for minor frequent poly (N) profiles, with their relative proportion in general increasing with the depth of sequencing (i.e., total number of reads counted). While this indicates that the depth of sequencing (or degree of sensitivity) applied by Illumina run is the key for detecting minor variants clones, it further corroborates that the applied downstream strategy seems not to have impact on the results.

Finally, the reliability of our results is additionally reinforced because: i) homopolymer-associated errors have been shown to be tremendously minimized using Illumina technology (126–129); ii) similar strategies have previously allowed determining that these kind of variable regions mediate relevant alterations in virulence-associated genes (126,127); and iii) for instance, 15 *T. pallidum* poly (G/C) tracts were recently found to be consistently conserved across multiple DNA samples (113), which excludes Illumina bias and constitutes a good proof of principle for the approach that was followed in the present study.

3.2 Allelic variation among strains

Based on recent studies at INSA on adaptation and phase variation mechanisms in *C. trachomatis* (104) we selected two promising targets, the cytotoxin gene (CT166) (Figure 19) and the *lpxC* gene (CT533) which codes for an enzyme involved in lipid A biosynthesis (Figure 20), to get insight on their profile of intra-patient variation.

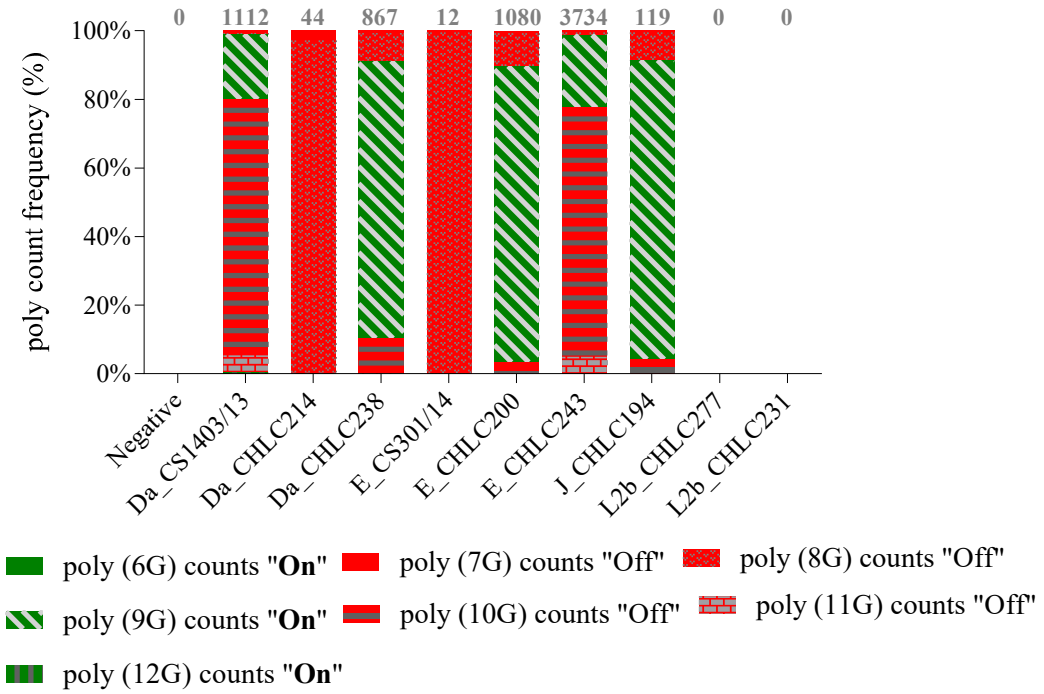

Figure 19 – Profile of the CT166 homopolymeric tract in several *C. trachomatis* strains. The figure shows the relative frequency of each poly (N) count retrieved from DNA samples directly obtained from *C. trachomatis*-positive clinical specimens. As indicated, depending on base count, the cytotoxin protein CT166 is predicted to be functional (ON) or truncated "OFF". On top of each graph bar, it is shown the total number of poly (N)-containing reads counted for each strain (in grey).

The cytotoxin gene evidences a well-known presence/absence profile among *C. trachomatis* strains (94), being present in genital genotypes, partially in ocular genotypes and absent in LGV genotypes. In this study, we evidenced that beyond the variable presence depending on strain type, this gene shows intra-strain variability *in vivo*. On one hand, this finding corroborates and scales-up previous observations (104) pointing that CT166 might be activated/inactivated by *C. trachomatis* in the context of human infection. On the other hand, together with other factors (inter-genotype variable presence and location in the plasticity zone), it also sustains that this variable gene is potentially associated with virulence and tissue tropism, with the variable poly (N) being likely the key for this role. Of note, our experiments concerning this gene (Figure 19), included a

negative control to ensure the absence of false positives caused by contamination, and also an L2b strain that functioned as an internal control, as it lacks the cytotoxin gene (94). Both controls revealed negative (no reads), technically validating the work.

Figure 20, represents results for CT533 (*lpxC* gene). In these experiments a negative control was also included for excluding false positive results due to contamination.

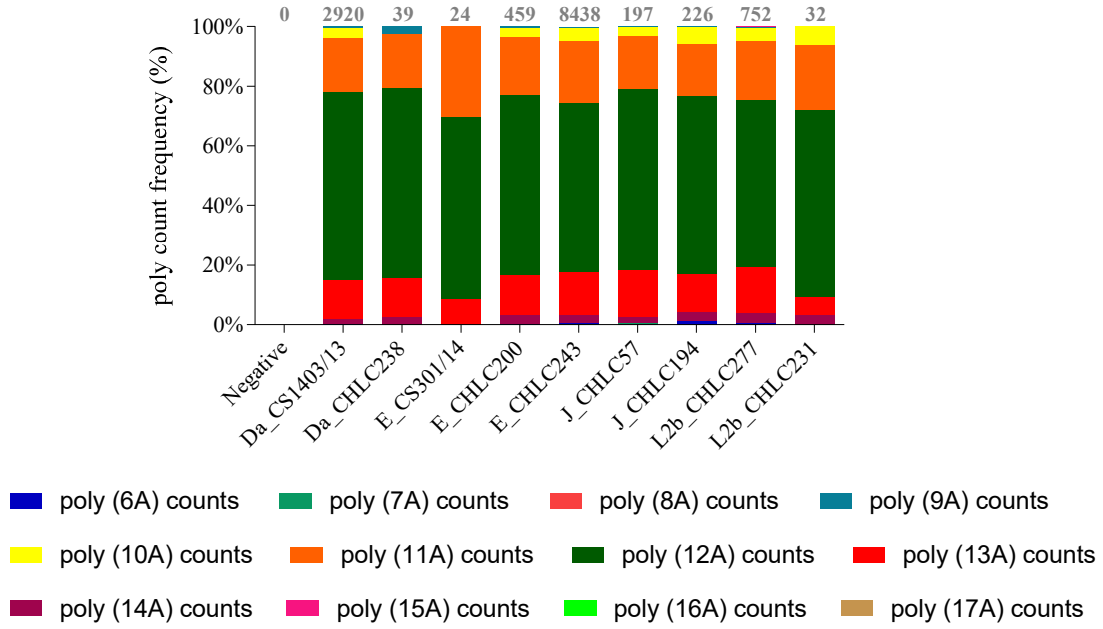

Figure 20 – Profile of the homopolymeric tract affecting the regulatory region of CT533 for several strains. The graph shows the relative frequency of each poly (N) count retrieved from DNA samples directly obtained from *C. trachomatis*-positive clinical specimens. On top of each graph bar, it is shown the total number of poly (N)-containing reads counted for each strain (in grey).

The homopolymeric tract located in putative regulatory region of the gene CT533 was shown to decrease in length (from 12 to 11 base counts) with increasing number of passaging cycles of *C. trachomatis ompA*-genotype E strain in cell culture (104). In the present study, the most abundant poly (N) count was the same (12 'A') after assessing several DNA samples directly obtained from *C. trachomatis*-positive clinical specimens using the amplicon-based approach. As such, considering that the major count of 12 'A' was also the most frequent count in early passages (104), we may wonder whether the decrease in length of the homopolymeric tract was related to adaptation to an in vitro life

style, or whether the 12bp homopolymeric tract is particularly advantageous during human infection. In fact, CT533 codifies for an enzyme related to the biosynthesis of lipid A, which is a component of the chlamydial LPS, an essential constituent of gram-negative bacteria cell wall (33); so, changes in its gene expression may affect *C. trachomatis* viability and ability to generate progeny. Still a large survey enrolling multiple samples is needed to further corroborate either one or both hypothesis, as well as to confirm/exclude that this poly (N) actually mediates phase variation through reversible switching during human infections. Finally, it is noteworthy that the CT533 homopolymeric tract is the longest of our study. So, we cannot exclude that its large size may potentiate PCR-derived errors, although as state above, no effect was observed when increasing the number of PCR cycles.

Despite the emergence of reports pointing that intra-strain variation of homopolymeric length may putatively drive different virulence characteristics of *C. trachomatis,* by phase variation, we speculate that other biological features (like tissue tropism) may also imply distinct sizes of homopolymeric tracts. In fact, from our preliminary inspection of dozens of *C. trachomatis* homopolymeric tracts, we have detected several presenting distinct lengths for strains displaying different tissue tropism (data not shown). As such, in order to assess the applicability of our amplicon-based approach to identify tissue tropism related poly (Ns) size of each target gene, we gathered the results obtained for a genital and a LGV strain (Figure 21) as a mean to identify/confirm tropism-specific differences.

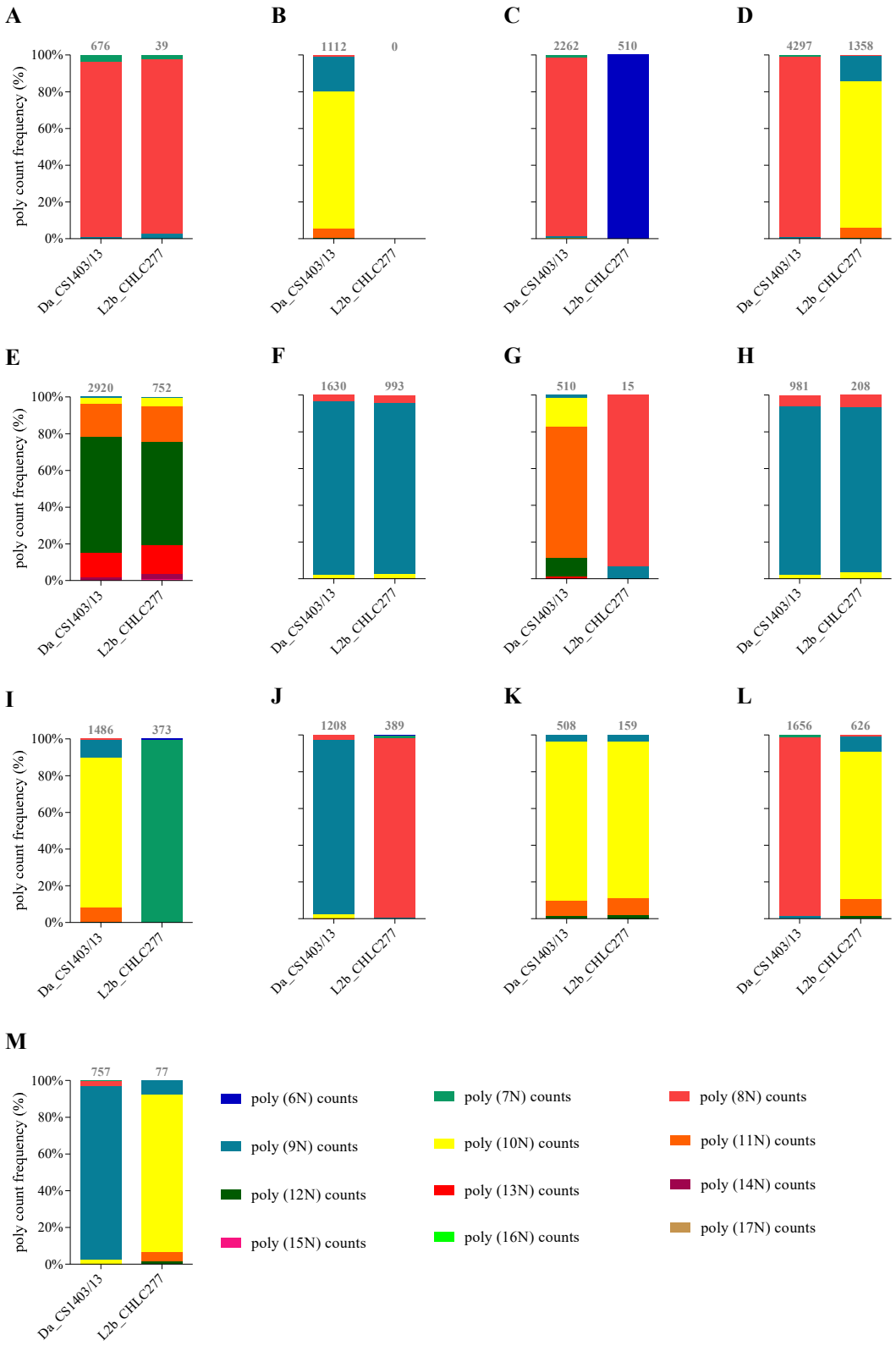

Figure 21 – Analysis of tropism-specific homopolymeric lengths. Each panel shows the result for a genital and LGV strain. On the top of each bar, it is presented the total number of poly (N)-containing reads counted for each strain (in grey). Panel A shows the result for the poly (N) likely affecting the gene CT085,

panel B the result for CT166, panel C the result for CT226, panel D the result for CT326, panel E the result for CT533, panel F the result for CT541, panel G the result for CT561, panel H the result for CT572, panel I the result for CT605, panel J the result for CT694, panel K the result for CT801, panel L the result for CT823 and panel M the result for CT824.

As expected, considering our preliminary survey of multiple *C. trachomatis* genome sequences, some homopolymeric tracts revealed no differences in length between the two strains, whereas others exhibited great inter-strain differences. While our results indicate that the latter may be involved in tissue tropism differences, it further points out that our amplicon-based approach can indeed be straightforwardly applied to large DNA sample collections in a near future, to confirm the association between tissue tropism and poly (Ns) lengths. This is relevant, as some of the targeted genes may mediate important bacterial functions/phenotypic differences. For instance, CT226 codes for an inclusion membrane protein (130), CT326 has been related to rectal tissue tropism (131), CT561 is part of the secretion apparatus (23), CT694 codes for a protein that acts as substrate for the *C. trachomatis* secretion system and an early effector protein (132), CT823 encodes a serine protease that interacts with the host cell by manipulating the signal pathways (133), CT824 a zinc dependent protease, and finally CT166 encodes a cytotoxin which is, as stated above, absent or present in LGV or genital strains, respectively.

Finally, we highlight some target genes that presented high intra- and/or inter-strain variability, as they can be considered of interest for further studies (Figure 22).

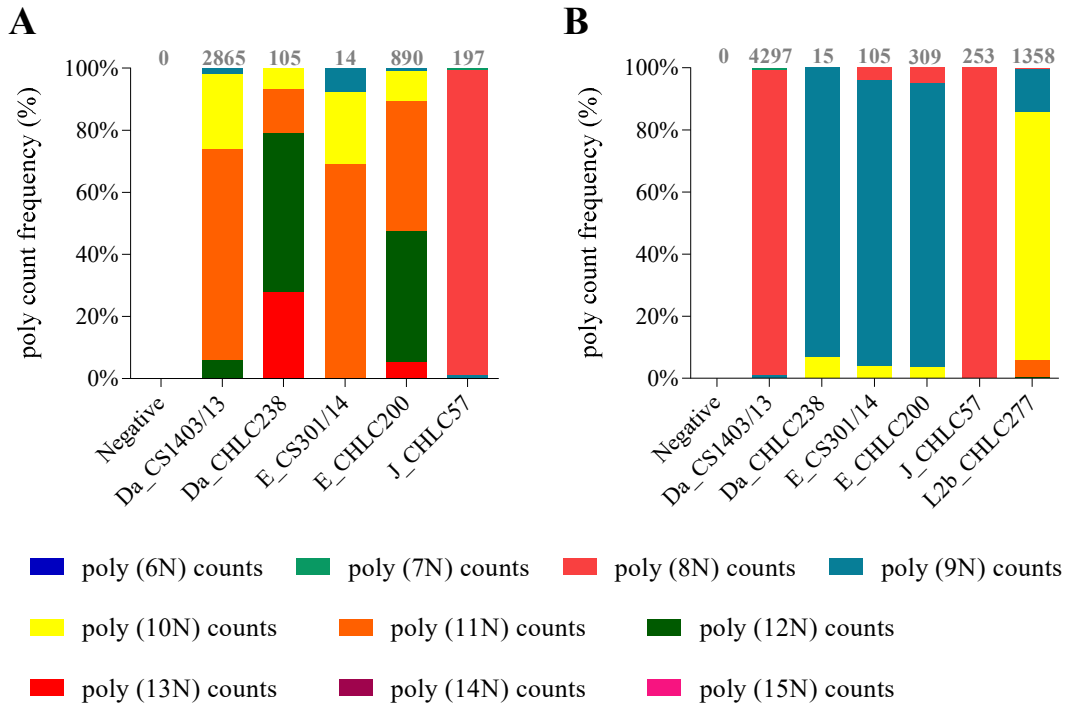

Figure 22 – Intra-strain homopolymeric tracts variability. Each panel shows the results for several strains. On top of each column it is shown the number of reads (in grey). Panel A shows the results for gene CT259 and panel B for gene CT326.

Beyond the cytotoxin-encoding gene (CT166) (Figure 19), we highlight the homopolymeric tract affecting CT259, a protein phosphatase associated to interact with the host through the signaling pathways (134), and CT326, as this hypothetical protein, as mentioned above, has been suggested to correlate to tissue tropism (131).

Taking all together, this preliminary study focused on implementing a technique that promotes a new insight on the genetic implications of the heterogeneity of homopolymeric tracts, potentially driving to phase variation in *C. trachomatis*. In fact, by applying this methodology to more strains and more targets, we may open further research lines, that should contribute to a better understanding of the pathogenesis of such an important microorganism as *C. trachomatis*.

IV. Conclusion

The epidemiology of *C. trachomatis* is unknown in Portugal. In 2014, the Portuguese health authority included this infection in the list of obligatory reported diseases; however, no screening studies have been implemented and *C. trachomatis* testing is not supported by the national health system; thus, *C. trachomatis* testing is not often required and the knowledge about its epidemiology did not improve. In such a scenario, the knowledge about the *C. trachomatis ompA*-genotypes that infect the Portuguese population is poor. One of the purposes of the present work was to contribute to fulfill that gap of information by characterizing the frequency of circulating *C. trachomatis ompA*-genotypes within a collection of *C. trachomatis* Portuguese specimens. The importance of typing relies on prescribing the adequate therapeutic (for instance doxycycline instead of azithromycin in the case of LGV) and to evaluate the evolution of circulating chlamydial types, namely in relation to clinical data, sexual habits, age, or other factors that may reveal important to prevent *C. trachomatis* acquisition and long term sequelae. In fact, *C. trachomatis* infections are easily treatable and their high frequency worldwide is due to their asymptomatic character, justifying the implementation of *C. trachomatis* screening programs. In our study, the most frequent *C. trachomatis ompA*-genotypes were E and F; of note L2 strains were also frequent, due to the collection characteristics (many positive samples from MSM).

Regarding LGV, the need for rapid L1, L2 and L3 identification is crucial for therapeutic management of patients and to notify the case to health authorities. The traditional *ompA*-genotyping system does not provide results as fast as necessary. Thus, we contributed for the implementation of a protocol that would greatly reduce time. This task could not be completed because of time and budget limitations (acquisition of probes was required) but results were promising and it should provide the expected specific and fast results in a near future.

The number of *ompA*-variants was an interesting finding, and we could detect some new variants that time will tell if they will be fixed by *C. trachomatis* strains

circulating in Portugal, as it seems to be the case of variants that are being identified in the host laboratory since 1991.

Through the present thesis, we were given the opportunity to propagate *C. trachomatis* in cell culture, a methodology no longer used for diagnosis (eventually it may be required for forensic reasons, namely suspicion of sexual assault). This technique should be useful in the future for research studies requiring cell culture or *C. trachomatis* propagation.

The knowledge of *C. trachomatis* genome and diversity has been extremely enriched with the recent wide use of techniques such whole genome sequencing, which allow us a straightforward characterization of the inter-strain genetic diversity. In fact, it rapidly provides the opportunity to compare the genomes between *ompA*-genotypes, and thus identifying the genetic traits that characterize each *ompA*-genotype and also each biovar. Still, increasing evidences has been showing that efforts to fully understand the virulence and adaptation of each chlamydial strain should also be focused on the evaluation of the *in vivo* intra-strain variability. This encouraged us to perform a study aiming at identifying phase variation mechanisms potentially mediated by homopolymeric tracts. Besides the implementation of a protocol to study these mechanisms directly from the clinical samples, we were also able to identify some potential targets of phase variation events, likely driving bacterial adaptation and virulence in the context of the human infection, as exemplified by the interesting case of the cytotoxin gene (CT166). *C. trachomatis* genes of unknown function, many coding for hypothetical proteins exclusive of the bacteria, and other gene targets should be evaluated regarding their homopolymeric tracts, among an enlarged collection of *C. trachomatis* clinical strains, as this approach seems a promising research area.

New genomic, proteomic and microscopy tools should contribute to better understand bacterial-host adaptation and would surely contribute for a better understanding of chlamydial induced pathogenesis, a required knowledge to implement more precise preventive measures, namely the development of a vaccine.

V. References

1. World Health Organization. Global Strategy for the Prevention and Control of Sexually Transmitted Infections: 2006-2015. World Heal Organ. 2007;27(10):61.

2. World Health Organization. Global Health Sector Strategy on Sexually Transmitted Infections 2016-2021. World Heal Organ. 2016;

3. Stephens RS, Myers G, Eppinger M, Bavoil PM. Divergence without difference: Phylogenetics and taxonomy of *Chlamydia* resolved. FEMS Immunol Med Microbiol. 2009;55(2):115–9.

4. Sachse K, Bavoil PM, Kaltenboeck B, Stephens RS, Kuo C-C, Rosselló-Móra R, et al. Emendation of the family *Chlamydiaceae*: Proposal of a single genus, *Chlamydia*, to include all currently recognized species. Syst Appl Microbiol. Elsevier GmbH.; 2015 Mar;38(2):99–103.

5. Sachse K, Laroucau K, Riege K, Wehner S, Dilcher M, Creasy HH, et al. Evidence for the existence of two new members of the family *Chlamydiaceae* and proposal of *Chlamydia avium* sp. nov. and *Chlamydia gallinacea* sp. nov. Syst Appl Microbiol. Elsevier GmbH.; 2014 Mar;37(2):79–88.

6. Bébéar C, de Barbeyrac B. Genital *Chlamydia trachomatis* infections. Clin Microbiol Infect. 2009 Jan;15(1):4–10.

7. Roulis E, Polkinghorne A, Timms P. *Chlamydia pneumoniae*: modern insights into an ancient pathogen. Trends Microbiol. Elsevier Ltd; 2013 Mar;21(3):120–8.

8. Rizzo A, Domenico M Di, Carratelli CR, Paolillo R. The Role of *Chlamydia* and *Chlamydophila* Infections in Reactive Arthritis. Intern Med. 2012;51(1):113–7.

9. Deniset JF, Cheung PKM, Dibrov E, Lee K, Steigerwald S, Pierce GN. *Chlamydophila pneumoniae* Infection Leads to Smooth Muscle Cell Proliferation and Thickening in the Coronary Artery without Contributions from a Host Immune Response. Am J Pathol. 2010 Feb;176(2):1028–37.

10. Maheshwari P, Eslick GD. Bacterial infection and Alzheimer's disease: a meta-analysis. J Alzheimers Dis. 2015;43(3):957–66.

11. Radomski N, Einenkel R, Müller A, Knittler MR. *Chlamydia*-host cell interaction not only from a bird's eye view: some lessons from *Chlamydia psittaci*. FEBS Lett. 2016 Nov;590(21):3920–40.

12. Abdelrahman YM, Belland RJ. The chlamydial developmental cycle. FEMS Microbiol Rev. 2005 Nov;29(5):949–59.

13. Elwell C, Mirrashidi K, Engel J. *Chlamydia* cell biology and pathogenesis. Nat Rev Microbiol. Nature Publishing Group; 2016 Apr 25;14(6):385–400.

14. Nans A, Ford C, Hayward RD. Host-pathogen reorganisation during host cell entry by *Chlamydia trachomatis*. Microbes Infect. Elsevier Masson SAS; 2015 Nov;17(11–12):727–31.

15. Abdelrahman Y, Ouellette SP, Belland RJ, Cox J V. Polarized Cell Division of *Chlamydia trachomatis*. Sassetti CM, editor. PLoS Pathog. 2016 Aug 9;12(8):e1005822.

16. Beatty WL, Morrison RP, Byrne GI. Persistent *chlamydiae*: from cell culture to a paradigm for chlamydial pathogenesis. Microbiol Rev. 1994 Dec;58(4):686–99.

17. Karin D Everett. www.chlamydiae.com [Internet]. Available from: http://chlamydiae.com/twiki/bin/view/Cell_Biology/GrowthCycle

18. Bavoil P, Tan M, editors. Intracellular Pathogens I: *Chlamydiales*. Intracellular Pathogens I: *Chlamydiales*. American Society of Microbiology; 2012.

19. Nikaido H. Molecular basis of bacterial outer membrane permeability revisited. Microbiol Mol Biol Rev. 2003 Dec 1;67(4):593–656.

20. Hsia RC, Pannekoek Y, Ingerowski E, Bavoil PM. Type III secretion genes identify a putative virulence locus of *Chlamydia*. Mol Microbiol. 1997 Jul 1;25(2):351–9.

21. Dai W, Li Z. Conserved type III secretion system exerts important roles in *Chlamydia trachomatis*. Int J Clin Exp Pathol. 2014;7(9):5404–14.

22. Hueck CJ. Type III protein secretion systems in bacterial pathogens of animals and plants. Microbiol Mol Biol Rev. 1998 Jun;62(2):379–433.

23. Betts-Hampikian HJ, Fields KA. The *Chlamydial* Type III Secretion Mechanism: Revealing Cracks in a Tough Nut. Front Microbiol. 2010;1(OCT):114.

24. Liu X, Afrane M, Clemmer DE, Zhong G, Nelson DE. Identification of *Chlamydia trachomatis* outer membrane complex proteins by differential proteomics. J Bacteriol. 2010 Jun 1;192(11):2852–60.

25. Caldwell HD, Kromhout J, Schachter J. Purification and partial characterization of the major outer membrane protein of *Chlamydia trachomatis*. Infect Immun. 1981 Mar;31(3):1161–76.

26. Hatch TP, Allan I, Pearce JH. Structural and polypeptide differences between envelopes of infective and reproductive life cycle forms of *Chlamydia* spp. J Bacteriol. 1984 Jan;157(1):13–20.

27. Nunes A, Borrego MJ, Nunes B, Florindo C, Gomes JP. Evolutionary dynamics of ompA, the gene encoding the *Chlamydia trachomatis* key antigen. J Bacteriol. 2009 Dec 1;191(23):7182–92.

28. Su H, Watkins NG, Zhang YX, Caldwell HD. *Chlamydia trachomatis*-host cell interactions: role of the chlamydial major outer membrane protein as an adhesin. Infect Immun. 1990 Apr;58(4):1017–25.

29. Bavoil P, Ohlin A, Schachter J. Role of disulfide bonding in outer membrane structure and permeability in *Chlamydia trachomatis*. Infect Immun. 1984 May;44(2):479–85.

30. Kubo A, Stephens RS. Characterization and functional analysis of PorB, a *Chlamydia* porin and neutralizing target. Mol Microbiol. 2000 Nov;38(4):772–80.

31. Gomes JP, Nunes A, Bruno WJ, Borrego MJ, Florindo C, Dean D. Polymorphisms in the Nine Polymorphic Membrane Proteins of *Chlamydia trachomatis* across All Serovars: Evidence for Serovar Da Recombination and Correlation with Tissue Tropism. J Bacteriol. 2006 Jan 1;188(1):275–86.

32. Henderson IR, Lam AC. Polymorphic proteins of *Chlamydia* spp. – autotransporters beyond the Proteobacteria. Trends Microbiol. 2001 Dec;9(12):573–8.

33. Nguyen BD, Cunningham D, Liang X, Chen X, Toone EJ, Raetz CRH, et al. Lipooligosaccharide is required for the generation of infectious elementary bodies in *Chlamydia trachomatis*. Proc Natl Acad Sci. 2011 Jun 21;108(25):10284–9.

34. Fadel S, Eley A. Is lipopolysaccharide a factor in infectivity of *Chlamydia trachomatis*? J Med Microbiol. 2008 Mar 1;57(3):261–6.

35. Liechti GW, Kuru E, Hall E, Kalinda A, Brun Y V, VanNieuwenhze M, et al. A new metabolic cell-wall labelling method reveals peptidoglycan in *Chlamydia trachomatis*. Nature. Nature Publishing Group; 2013 Dec 11;506(7489):507–10.

36. Skipp P, Robinson J, O'Connor CD, Clarke IN. Shotgun proteomic analysis of *Chlamydia trachomatis*. Proteomics. 2005 Apr;5(6):1558–73.

37. Aistleitner K, Anrather D, Schott T, Klose J, Bright M, Ammerer G, et al. Conserved features and major differences in the outer membrane protein composition of *chlamydiae*. Environ Microbiol. 2015 Apr;17(4):1397–413.

38. Jewett TJ, Fischer ER, Mead DJ, Hackstadt T. Chlamydial TARP is a bacterial nucleator of actin. Proc Natl Acad Sci. 2006 Oct 17;103(42):15599–604.

39. Lutter EI, Martens C, Hackstadt T. Evolution and Conservation of Predicted Inclusion Membrane Proteins in *Chlamydiae*. 2012;2012.

40. Nicholson TL, Olinger L, Chong K, Schoolnik G, Stephens RS. Global Stage-Specific Gene Regulation during the Developmental Cycle of *Chlamydia trachomatis*. J Bacteriol. 2003 May 15;185(10):3179–89.

41. Yang C, Starr T, Song L, Carlson JH, Sturdevant GL, Beare PA, et al. Chlamydial lytic exit from host cells is plasmid regulated. MBio. 2015;6(6):1–9.

42. Todd WJ, Caldwell HD. The Interaction of *Chlamydia trachomatis* with Host Cells: Ultrastructural Studies of the Mechanism of Release of a Biovar II Strain from HeLa 229 Cells. J Infect Dis. 1985 Jun 1;151(6):1037–44.

43. Hybiske K, Stephens RS. Mechanisms of host cell exit by the intracellular bacterium *Chlamydia*. Proc Natl Acad Sci. 2007 Jul 3;104(27):11430–5.

44. Zuck M, Ellis T, Venida A, Hybiske K. Extrusions are phagocytosed and promote *Chlamydia* survival within macrophages. Cell Microbiol. 2016 Oct 14;

45. Schust DJ, Ibana JA, Buckner LR, Ficarra M, Sugimoto J, Amedee AM, et al. Potential mechanisms for increased HIV-1 transmission across the endocervical epithelium during *C. trachomatis* infection. Curr HIV Res. 2012 Apr;10(3):218–27.

46. Wang S -p., Kuo C -c., Barnes RC, Stephens RS, Grayston JT. Immunotyping of *Chlamydia trachomatis* with Monoclonal Antibodies. J Infect Dis. 1985 Oct 1;152(4):791–800.

47. Byrne GI. *Chlamydia trachomatis* Strains and Virulence: Rethinking Links to Infection Prevalence and Disease Severity. J Infect Dis. 2010 Jun 15;201(S2):126–33.

48. Yuan Y, Zhang YX, Watkins NG, Caldwell HD. Nucleotide and deduced amino acid sequences for the four variable domains of the major outer membrane proteins of the 15 *Chlamydia trachomatis* serovars. Infect Immun. 1989 Apr;57(4):1040–9.

49. Stephens RS, Sanchez-Pescador R, Wagar EA, Inouye C, Urdea MS. Diversity of *Chlamydia trachomatis* major outer membrane protein genes. J Bacteriol. 1987 Sep 1;169(9):3879–85.

50. Saitou N. and Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;(4):406–25.

51. Kimura M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 1980;(16):111–20.

52. Kumar S. SG and TK. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;(33):1870–4.

53. Brunelle BW, Sensabaugh GF. The ompA Gene in *Chlamydia trachomatis* Differs in Phylogeny and Rate of Evolution from Other Regions of the Genome. Infect Immun. 2006 Jan 1;74(1):578–85.

54. Taylor HR, Burton MJ, Haddad D, West S, Wright H. Trachoma. Lancet. 2014 Dec 13;384(9960):2142–52.

55. World Health Organization. Accelerating Work To Overcome the Global Impact of Neglected Tropical Diseases a Roadmap for Implementation. World Heal Organ. 2012;

56. Van de Laar MJ. Morré SA. *Chlamydia*: a major challenge for public health. Eurosurveillance [Internet]. 2007;12(10). Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=735

57. Cunningham KA, Beagley KW. Male Genital Tract Chlamydial Infection: Implications for Pathology and Infertility. Biol Reprod. 2008 Aug 1;79(2):180–9.

58. Mackern-Oberti JP, Motrich RD, Breser ML, Sánchez LR, Cuffini C, Rivero VE. *Chlamydia trachomatis* infection of the male genital tract: an update. J Reprod Immunol. Elsevier Ireland Ltd; 2013 Nov;100(1):37–53.

59. Price MJ, Ades A, Soldan K, Welton NJ, Macleod J, Simms I, et al. The natural history of *Chlamydia trachomatis* infection in women: a multi-parameter evidence synthesis. Health Technol Assess (Rockv). 2016 Mar;20(22):1–250.

60. STENBERG K. PERSISTENT NEONATAL CHLAMYDIAL INFECTION IN A 6-YEAR-OLD GIRL. Lancet. 1986 Nov;328(8518):1278–9.

61. Mabey D, Peeling RW. Lymphogranuloma venereum. Sex Transm Infect. 2002 Apr 1;78(2):90–2.

62. White J a. Manifestations and management of lymphogranuloma venereum. Curr Opin Infect Dis. 2009 Feb;22(1):57–66.

63. van Hal SJ, Hillman R, Stark DJ, Harkness JL, Marriott D. Lymphogranuloma venereum: an emerging anorectal disease in Australia. Med J Aust. 2007 Sep 3;187(5):309–10.

64. De Vries HJC, Zingoni A, Kreuter A, Moi H, White JA. 2013 European guideline on the management of lymphogranuloma venereum. J Eur Acad Dermatology Venereol. 2015;29(1):1–6.

65. Stary G, Stary A. Lymphogranuloma venereum outbreak in Europe. JDDG. 2008 May 21;2(1):935–9.

66. Savage EJ, Laar MJ Van De, Gallay A, Sande M Van Der, Hamouda O, Sasse A, et al. Lymphogranuloma venereum in Europe, 2003-2008. 2009;1–5. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19428

67. Herida M, Michel A, Goulet V, Janier M, Sednaoui P, Dupin N, et al. L'épidémiologie des infections sexuellement transmissibles en France. Médecine Mal Infect. 2005 May;35(5):281–9.

68. Ward H, Martin I, Macdonald N, Alexander S, Simms I, Fenton K, et al. Lymphogranuloma Venereum in the United Kingdom. Clin Infect Dis. 2007 Jan;44(1):26–32.

69. Nieuwenhuis RF, Ossewaarde JM, Götz HM, Dees J, Thio HB, Thomeer MGJ, et al. Resurgence of Lymphogranuloma Venereum in Western Europe: An Outbreak of *Chlamydia trachomatis* Serovar L 2 Proctitis in The Netherlands among Men Who Have Sex with Men. Clin Infect Dis. 2004 Oct;39(7):996–1003.

70. Kropp RY. Emergence of lymphogranuloma venereum in Canada. Can Med Assoc J. 2005 Jun 21;172(13):1674–6.

71. Gomes JP, Nunes A, Florindo C, Ferreira MA, Santo I, Azevedo J, et al. Lymphogranuloma Venereum in Portugal. Sex Transm Dis. 2009 Feb;36(2):88–91.

72. Castro R, Baptista T, Vale A, Nunes H, Prieto E, Araújo C, et al. Lymphogranuloma venereum serovar L2b in Portugal. Int J STD AIDS. 2010 Apr 1;21(4):265–6.

73. Ripa T. Nilsson T. A variant of *Chlamydia trachomatis* with deletion in cryptic plasmid: implications for use of PCR diagnostic test. Eurosurveillance [Internet]. 2006;11(45). Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=3076

74. Unemo M, Clarke IN. The Swedish new variant of *Chlamydia trachomatis*. Curr Opin Infect Dis. 2011 Feb;24(1):62–9.

75. Jurstrand M, Fredlund H, Unemo M. The new variant of *Chlamydia trachomatis* was present as early as 2003 in Örebro County, Sweden, but remained undetected until 2006. Sex Transm Infect. 2013 Nov;89(7):607–8.

76. Christerson L, Bom RJM, Bruisten SM, Yass R, Hardick J, Bratt G, et al. *Chlamydia trachomatis* Strains Show Specific Clustering for Men Who Have Sex with Men Compared to Heterosexual Populations in Sweden, the Netherlands, and the United States. J Clin Microbiol. 2012 Nov 1;50(11):3548–55.

77. Lanjouw E, Ouburg S, de Vries HJ, Stary A, Radcliffe K, Unemo M. Background review for the "2015 European guideline on the management of *Chlamydia trachomatis* infections". Int J STD AIDS. 2015 Nov 24;0(0):1–25.

78. World Health Organization. WHO guidelines for the treatment of *Chlamydia trachomatis*. World Heal Organ. 2016;

79. World Health Organization. Report on global sexually transmitted infection surveillance 2013. WHO Libr Cat Data. 2014;1–54.

80. Centers for Disease Control and Prevention. Sexually Transmitted Diseases Treatment Guidelines. Vol. 64. 2015.

81. Magliano E. The microbiological diagnosis of *Chlamydia trachomatis* infections: milestones from a centenary history. Microbiol Medica. 2016 Jul 1;31(2):63–5.

82. World Health Organization. Laboratory diagnosis of sexually transmitted infections , including human immunodeficiency virus. World Heal Organ. 2013;

83. Persson K. The role of serology, antibiotic susceptibility testing and serovar determination in genital chlamydial infections. Best Pract Res Clin Obstet Gynaecol. 2002 Dec;16(6):801–14.

84. Pedersen LN, Herrmann B, MÃ¸ller JK. Typing *Chlamydia trachomatis* : from egg yolk to nanotechnology. FEMS Immunol Med Microbiol. 2009 Mar;55(2):120–30.

85. Dubuis O, Gorgievski-Hrisoho M, Germann D, Matter L. Evaluation of 2-SP transport medium for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* by

two automated amplification systems and culture for *chlamydia*. J Clin Pathol. 1997 Nov 1;50(11):947–50.

86. Parra-Sánchez M, Palomares JC, Bernal S, González MT, Sivianes N, Pérez L, et al. Evaluation of the cobas 4800 CT/NG Test for detecting *Chlamydia trachomatis* and *Neisseria gonorrhoeae* DNA in urogenital swabs and urine specimens. Diagn Microbiol Infect Dis. Elsevier Inc.; 2012 Dec;74(4):338–42.

87. Zomorodipour A, Andersson SGE. Obligate intracellular parasites: *Rickettsia prowazekii* and *Chlamydia trachomatis*. FEBS Lett. 1999;452(1–2):11–5.

88. Andersson JO, Andersson SG. Insights into the evolutionary process of genome degradation. Curr Opin Genet Dev. 1999 Dec;9(6):664–71.

89. Stephens RS. Genome Sequence of an Obligate Intracellular Pathogen of Humans: *Chlamydia trachomatis*. Science (80-). 1998 Oct 23;282(5389):754–9.

90. Clarke IN. Evolution of *Chlamydia trachomatis*. Ann N Y Acad Sci. 2011;1230:11–8.

91. Nunes A, Gomes JP. Evolution, phylogeny, and molecular epidemiology of *Chlamydia*. Infect Genet Evol. Elsevier B.V.; 2014 Apr;23(February):49–64.

92. Comanducci M, Ricci S, Cevenini R, Ratti G. Diversity of the *Chlamydia trachomatis* common plasmid in biovars with different pathogenicity. Plasmid. 1990 Mar;23(2):149–54.

93. Caldwell HD, Wood H, Crane D, Bailey R, Jones RB, Mabey D, et al. Polymorphisms in *Chlamydia trachomatis* tryptophan synthase genes differentiate between genital and ocular isolates. J Clin Invest. 2003 Jun 1;111(11):1757–69.

94. Carlson JH, Hughes S, Hogan D, Cieplak G, Sturdevant DE, McClarty G, et al. Polymorphisms in the *Chlamydia trachomatis* Cytotoxin Locus Associated with Ocular and Genital Isolates. Infect Immun. 2004 Dec 1;72(12):7063–72.

95. Belland RJ, Scidmore MA, Crane DD, Hogan DM, Whitmire W, McClarty G, et al. *Chlamydia trachomatis* cytotoxicity associated with complete and partial cytotoxin genes. Proc Natl Acad Sci. 2001 Nov 20;98(24):13984–9.

96. Carabeo R. Bacterial subversion of host actin dynamics at the plasma membrane. Cell Microbiol. 2011 Oct;13(10):1460–9.

97. Taylor LD, Nelson DE, Dorward DW, Whitmire WM, Caldwell HD. Biological Characterization of *Chlamydia trachomatis* Plasticity Zone MACPF Domain Family Protein CT153. 2010;78(6):2691–9.

98. Nunes A, Borrego MJ, Gomes JP. Genomic features beyond *Chlamydia trachomatis* phenotypes: What do we think we know? Infect Genet Evol. Elsevier B.V.; 2013 Jun;16:392–400.

99. Clifton DR, Fields KA, Grieshaber SS, Dooley CA, Fischer ER, Mead DJ, et al. From The Cover: A chlamydial type III translocated protein is tyrosine-phosphorylated at the site of entry and associated with recruitment of actin. Proc Natl Acad Sci. 2004 Jul 6;101(27):10166–71.

100. Lutter EI, Bonner C, Holland MJ, Suchland RJ, Stamm WE, Jewett TJ, et al. Phylogenetic Analysis of *Chlamydia trachomatis* Tarp and Correlation with Clinical Phenotype. Infect Immun. 2010 Sep 1;78(9):3678–88.

101. Lukácová M, Baumann M, Brade L, Mamat U, Brade H. Lipopolysaccharide smooth-rough phase variation in bacteria of the genus *Chlamydia*. Infect Immun. 1994 Jun;62(6):2270–6.

102. Read TD, Brunham RC, Shen C, Gill SR, Heidelberg JF, White O, et al. Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. Nucleic Acids Res. 2000 Mar 15;28(6):1397–406.

103. Abdelsamed H, Peters J, Byrne GI. Genetic variation in *Chlamydia trachomatis* and their hosts: impact on disease severity and tissue tropism. Future Microbiol. 2013 Sep;8(9):1129–46.

104. Borges V, Pinheiro M, Antelo M, Sampaio DA, Vieira L, Ferreira R, et al. *Chlamydia trachomatis* In Vivo to In Vitro Transition Reveals Mechanisms of Phase Variation and Down-Regulation of Virulence Factors. Ojcius DM, editor. PLoS One. 2015 Jul 24;10(7):e0133420.

105. Nunes A, Gomes JP, Karunakaran KP, Brunham RC. Bioinformatic analysis of *Chlamydia trachomatis* polymorphic membrane proteins PmpE, PmpF, PmpG and PmpH as potential vaccine antigens. PLoS One. 2015;10(7):1–16.

106. Salaün L, Snyder LAS, Saunders NJ. Adaptation by Phase Variation in Pathogenic Bacteria. In: Advances In Applied Microbiology. 2003. p. 263–301.

107. van Belkum a, Scherer S, van Alphen L, Verbrugh H. Short-sequence DNA repeats in prokaryotic genomes. Microbiol Mol Biol Rev. 1998;62(2):275–93.

108. Levinson G, Gutman G a. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol. 1987;4(3):203–21.

109. Lan J, Walboomers JM, Roosendaal R, van Doornum GJ, MacLaren DM, Meijer CJ, et al. Direct detection and genotyping of *Chlamydia trachomatis* in cervical scrapes by using polymerase chain reaction and restriction fragment length polymorphism analysis. J Clin Microbiol. 1993 May;31(5):1060–5.

110. Sequencing chemistry [Internet]. Available from: http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/dna-sequencing-fragment-analysis/overview-of-dna-sequencing/sequencing-chemistries.html

111. Szumilas M. Explaining odds ratios. J Can Acad Child Adolesc Psychiatry. 2010 Aug 18;19(3):227–9.

112. Möbius N, Brenneisen W, Schaeffer A, Henrich B. Protocol for the Rapid Detection of the Urogenital Tract Mollicutes and *Chlamydia* with Concomitant LGV-(sub)typing. In: MacKenzie CR, Henrich B, editors. Methods in molecular biology (Clifton, NJ). Totowa, NJ: Humana Press; 2012. p. 235–53. (Methods in Molecular Biology; vol. 965).

113. Pinto M, Borges V, Antelo M, Pinheiro M, Nunes A, Azevedo J, et al. Genome-scale analysis of the non-cultivable *Treponema pallidum* reveals extensive within-patient genetic variation. Nat Microbiol. Nature Publishing Group; 2016 Oct 17;2(October):16190.

114. Black CM, Marrazzo J, Johnson RE, Hook III EW, Jones RB, Green TA, et al. Head-to-Head Multicenter Comparison of DNA Probe and Nucleic Acid Amplification Tests for *Chlamydia trachomatis* Infection in Women Performed with an Improved Reference Standard. J Clin Microbiol. 2002 Oct 1;40(10):3757–63.

115. Sulaiman S, Chong PP, Mokhtarudin R, Lye MS, Wan Hassan WH. Comparison of nested and ELISA based polymerase chain reaction assays for detecting *Chlamydia trachomatis* in pregnant women with preterm complications. Trop Biomed. 2014 Mar;31(1):36–45.

116. Geisler WM, Lensing SY, Press CG, Hook EW. Spontaneous Resolution of Genital *Chlamydia trachomatis* Infection in Women and Protection from Reinfection. J Infect Dis. 2013 Jun 15;207(12):1850–6.

117. Peters RPH, Nijsten N, Mutsaers J, Jansen CL, Morré SA, van Leeuwen AP. Screening of oropharynx and anorectum increases prevalence of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* infection in female STD clinic visitors. Sex Transm Dis. 2011 Sep;38(9):783–7.

118. Borrego MJ, Gomes JP, Lefebvre JF, Eb F, Orfila J, Catry M a. Genotyping of Portuguese *Chlamydia trachomatis* urogenital isolates. Sex Transm Infect. 1997 Dec 1;73(6):561–3.

119. Gomes JP, Borrego MJ, Atik B, Santo I, Azevedo J, Brito de Sá A, et al. Correlating *Chlamydia trachomatis* infectious load with urogenital ecological success and disease pathogenesis. Microbes Infect. 2006 Jan;8(1):16–26.

120. Brasiliense DM, Borges BN, Ferreira WAS. Genotyping and prevalence of *Chlamydia trachomatis* infection among women in Belém, Pará, northern Brazil. J Infect Dev Ctries. 2016 Feb 28;10(2):237–50.

121. Taheri Beni B, Beni BT, Motamedi H, Ardakani MR. Genotyping of the prevalent *Chlamydia trachomatis* strains involved in cervical infections in women in Ahvaz, Iran. J Med Microbiol [Internet]. 2010 Sep 1;59(Pt 9):1023–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20508000

122. de Jesús De Haro-Cruz M, Deleón-Rodriguez I, Escobedo-Guerra MR, López-Hurtado M, Arteaga-Troncoso G, Ortiz-Ibarra FJ, et al. Genotyping of *Chlamydia trachomatis* from endocervical specimens of infertile Mexican women. Enferm Infecc Microbiol Clin. 2011 Feb;29(2):102–8.

123. Nunes A, Nogueira PJ, Borrego MJ, Gomes JP. *Chlamydia trachomatis* diversity viewed as a tissue-specific coevolutionary arms race. Genome Biol. 2008 Oct 23;9(10):R153.

124. Tanzer RJ, Hatch TP. Characterization of Outer Membrane Proteins in *Chlamydia trachomatis* LGV Serovar L2. J Bacteriol. 2001 Apr 15;183(8):2686–90.

125. Gomes JP, Bruno WJ, Nunes A, Santos N, Florindo C, Borrego MJ, et al. Evolution of *Chlamydia trachomatis* diversity occurs by widespread interstrain recombination involving hotspots. Genome Res. 2007;17(1):50–60.

126. Dohm JC, Lottaz C, Borodina T, Himmelbauer H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res. 2008 Sep 1;36(16):e105.

127. Jerome JP, Bell JA, Plovanich-Jones AE, Barrick JE, Brown CT, Mansfield LS. Standing genetic variation in contingency loci drives the rapid adaptation of *Campylobacter jejuni* to a novel host. Bereswill S, editor. PLoS One. 2011 Jan 24;6(1):e16399.

128. Loman NJ, Misra R V, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, et al. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol. 2012 May 22;30(5):434–9.

129. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics. BMC Genomics; 2012 Jul 24;13(1):341.

130. Sharma J, Zhong Y, Dong F, Piper JM, Wang G, Zhong G. Profiling of human antibody responses to *Chlamydia trachomatis* urogenital tract infection using microplates arrayed with 156 chlamydial fusion proteins. Infect Immun. 2006 Mar 1;74(3):1490–9.

131. Jeffrey BM, Suchland RJ, Quinn KL, Davidson JR, Stamm WE, Rockey DD. Genome sequencing of recent clinical *Chlamydia trachomatis* strains identifies loci associated with tissue tropism and regions of apparent recombination. Infect Immun. 2010 Jun 1;78(6):2544–53.

132. Hower S, Wolf K, Fields KA. Evidence that CT694 is a novel *Chlamydia trachomatis* T3S substrate capable of functioning during invasion or early cycle development. Mol Microbiol. 2009 Jun;72(6):1423–37.

133. Wu X, Lei L, Gong S, Chen D, Flores R, Zhong G. The chlamydial periplasmic stress response serine protease cHtrA is secreted into host cell cytosol. BMC Microbiol. 2011 Apr 28;11:87.

134. Verma A, Maurelli AT. Identification of two eukaryote-like serine/threonine kinases encoded by *Chlamydia trachomatis* serovar L2 and characterization of interacting partners of Pkn1. Infect Immun. 2003 Oct 1;71(10):5772–84.

VI. Annexes

Annex 1: Calculations for cell concentration

CELL SPLIT

1,2x105 HeLa 229 cells/cm2 (= per well)

(if we start late in the afternoon for next day morning, we can go up to 1,4 x 105 HeLa 229 cells / well)

96 well = 0,32 cm^2	3,84 x 104 HeLa 229 cells / well
48 well = 0,95 cm2	1,1 x 105 HeLa 229 cells / well
24 well = 1,9 cm2	2,3 x 105 HeLa 229 cells / well
6 well = 9,5 cm2	1,2 x 106 HeLa 229 cells / well

(these can be 'rutt round' numbers and not the arithmetical exact numbers)

Date: _____ seed HeLa229 (cell line name) cells in 1 (no of plates) x 24 (no. of wells/plate) well plate (s)

Calculate for 230000 (no. of cells per well) **/well x** 24 **well/plate x** 1 **,5 plates**

[when working with 96 well-plates and we just want 1 plate, we might consider only the necessary for, let's say, 120 wells, and the appropriate volume for those (100ul x 120 wells = 12ml)]

- Volume for each well: 1 ml
- Total of cells needed: 8280000
- Final volume (**cells + DMEM-10**) to distribute: 36

- Number of T 175 flasks: 2 (P____)(____% confluent)
- Volume of suspesion *per* flask: 10 ml

Cell counting

- Add **50ul** of **cell suspensio**n to 200ul of HB Hanks 1x >>> vortex (dilution 1:5)
- Add **50ul of the previous mix** to 50ul of Trypan Blue >>> vortex (dilution 1:2)
- Pipet 10ul of the previous mix (i.e., dilution 1:10) into each **hemocytometer** chamber
- Count 2 fields of each **hemocytometer** chamber

Number of cells (blue cells are not counted)

	Field 1	Field 2	Average
Chamber 1	55	49	52
Chamber2	50	57	53,5

Final average 52,75

Neubauer chamber factor	10000
Cell dilution for counting	10

Concentration: 5275000 **cells/ml in** 20 **ml of cell suspension**

Volume of cell suspension needed: 1,57 **ml of cell suspension**

Cell dispensing:

Add 1,56967 ml of cell suspension to 34,4303 ml of DMEM-10.

Dispense 1 ml per well across 1 x 24 -well plates

Annex 2: Oligonucleotides for Nested-PCR

Name	Sequence (5'→3')
NLO	ATGAAAAAACTCTTGAAATCG
NRO	CTCAACTGTAACTGCGTATTT
PCTM3	TCCTTGCAAGCTCTGCCTGTGGGGAATCCT
SERO2A	TTTCTAGATTTCATCTTGTT

Annex 3: Oligonucleotides for sequencing reaction

Name	Sequence (5'→3')
OMPA-1	TTATGATCGACGGAATTCT
SERO2A	TTTCTAGATTTCATCTTGTT

Annex 4: Oligonucleotides for subtyping LGV: Using sets of primers discriminatory of L1, L2 and L3

Name	Sequence (5'→3')
L2_ompA_F	CATGCGTATGGGTTACTATGGTGA
L2_ompA_R	TAAAGTCGCGCATCCACATTC
L1_ompA_R	TAAAGTTGCACATCCACATTC

Annex 5: Oligonucleotides for subtyping LGV: real time-PCR for *C. trachomatis ompA* and *pmpH*

Name (target)	Oligonucleotide (5'→ 3')
***C. trachomatis* (*ompA* gene)**	GGTTTCGGCGGAGATCCT
	AGTAACCMAYACGCATGCTGAT
LGV (*pmpH* gene)	CTGTGCCAACCTCATCATCAA
	AGACCCTTTCCGAGCATCACT
L1 (*ompA* gene)	CAGCATCTTTCAACTTAGTTGGGT TA
	AGCTCATATTTGGTACAGCATCCT T
L2 (*ompA* gene)	CAGCATCTTTCAACTTAGTTGGGTTAT
	TGATCTAAGCTCATATTTGGTACAAGCTTA
L3 (*ompA* gene)	CGCTTCCTTCAACTTAGTTGGATT
	TCAAAGCAGTGTTAGGAACAAGCT

Forward oligonucleotides are represented in grey, reverse oligonucleotides are represented in white. M – A or C, Y – T or C

Annex 6: Homopolymeric tracts [poly (N)] selected for study

Potential affected gene	Location of poly (N)	Poly (N) type *
CT042/glgX (glycogen hydrolase (debranching))	Possible regulatory region	G/C
CT043/slc1 (virulence factor)	Possible regulatory region	A/T
CT054/sucA (2-oxoglutarate dehydrogenase subunit E1)	Possible regulatory region	A/T
CT085 (hypothetical protein)	Coding region	G/C
CT134 (hypothetical protein)	Possible regulatory region	A/T
CT166 (cytotoxin)	Coding region	G/C
CT172 (hypothetical protein)	Possible regulatory region	G/C
CT226 (hypothetical protein)	Coding region	A/T
CT259 (PP2C phosphatase)	Possible regulatory region	G/C
CT326 (hypothetical protein)	Coding region	G/C
CT414/pmpC (polymorphic membrane protein C)	Possible regulatory region	A/T
CT445/gltX (glutamyl-tRNA synthetase)	Possible regulatory region	A/T
CT533/lpxC (LPS biosynthesis)	Possible regulatory region	A/T
CT541/mip (peptidyl-prolyl cis-trans isomerase)	Possible regulatory region	A/T
CT561 (type III secretion system protein)	Possible regulatory region	A/T
CT572/gspD (general secretion pathway protein D)	Coding region	G/C
CT590 (hypothetical protein)	Coding region	A/T
CT605 (hypothetical protein)	Possible regulatory region	A/T
CT694 (type III secretion system protein)	Possible regulatory region	A/T

CT739/ftsK (cell division protein FtsK)	Coding region	A/T
CT801/rpsF (30S ribosomal protein S6)	Coding region	A/T
CT823/htrA (DO serine protease)	Possible regulatory region	A/T
CT824 (zinc metalloprotease)	Possible regulatory region	A/T
CT871 /pmpG (polymorphic membrane protein G)	Coding region	G/C

***** The two hypotheses refer to the leading or lagging strand.

Annex 7: Clinical samples for homopolymeric tracts study

Name	Sample	Serovar
CS637_11_P6_control	Cervical swab	D
CS1025_11_P7_control	Cervical swab	E
CS1403_13	Cervical swab	Da
CS301_14	Urethral swab	E
CHLC57	Cervical swab	J
CHLC277	Anorectal swab	L2b
CHLC194	Urethral swab	J
CHLC243	Cervical swab	E
CHLC231	Anorectal swab	L2b
CHLC214	Anorectal swab	Da
CHLC238	Anorectal swab	Da
CHLC200	Cervical swab	E

Annex 8: Oligonucleotides for amplicon-based NGS

Name	Oligonucleotide (5'→ 3')
polyCT042	TGGCTCCTAGAGGTAAAGGGATAG
	CTGCGTGTTAGACATGAAGTATGC
polyCT043	CATAAAGGCGATCAGAGTGTTCTT
	ATCGCGGCTCTAATCATTTACTAA
polyCT054	CAACTGATAAGCCAAGCTCTGCTA
	TGGAGACACTTGCCCCTGTAAAT
polyCT085	AAGATTTTGCGTTGCCTCATCA
	CATCCGCTCTTTACCCAGACG
polyCT134	TTGTTATTTTGAAGGGTGGACAGT
	GTCTCTAAAGGGCAGCCAGCATA
polyCT166	CGAAGTTTATTACACAGCCAGTGA
	GTCCCGCTTCCGAATTTTTATCT
polyCT172	TCTCAACCAAATTCGGTAGAGGT
	GCTCCGGCTATTTTGTTTAGG
polyCT226	TATACAGGCAGAACCCAGTGTCAT
	TCTTTTGGTTGGGTAGATTAGGTA
polyCT259	ATTCCGTCTGCAATGGCAATGA
	AAGGGAGGCAACTTTTGATACCC
polyCT326	TTCGCAGCAGAATCAAGACTCA
	AGGGGATTGGGTCGTTTAGAAG
polyCT414	AGAAAACCCCTCTTACATA
	ATCAGCAGCTCGATACTCT
polyCT445	CATTCTTGGCGACTTTAACAAAAC
	CTACACGCATTGGTGCTATGAAAG
polyCT533	CACCCCGGAATAGCATACCTTAC

	TCCCTGGGAATCGAGAACTTG
polyCT541	GAACCGCCTCCGTTATCACATC
	TGCCGATTCAATTAAAAGAACTGG
polyCT561	TATTTTCGAGAGCGCTTAGAACAC
	ACCGCAGTATAGGCGTCTGG
polyCT572	GGATGATTCCAAAGGGCATAGTCT
	TAGGGGGAGGATCTTAGGTGAAAA
polyCT590	TACATGCCGTACGGTTACTGTCTC
	CCTTACTGGAGGAGGTTTTGTGT
polyCT605	GCTATCATCCTGCCCATACTCTC
	GCTGCAAATAAGCCTTTCTGAATA
polyCT694	TCAAAGCTAAATATCTGCGTATGC
	CCTCCGCCGAAGCAATAACTTTTA
polyCT739	GCGTTCCAACATGCGTAAAAT
	GTTGCGCAGTATTATTGTCTATGA
polyCT801	CTGAAGGGGTTGCTGATTATGTAT
	TAAAGCCTTACGTCTAGCGTCTTC
polyCT823	ATGCTGCATATCTGCTTGTTAT
	CAATAGCCTGGTTCCCTGTTTTAG
polyCT824	CCTGGCCTTATATTTTAGGAAGC
	ATCCCGAATTGGGTAACTCTCAGA
polyCT871	TGCAAACGTCTTTCCATAAGTTCT
	GTCCGTATGTTCTCGAAAGTCAAC

Forward oligonucleotides are represented in grey, reverse oligonucleotides are represented in white

Annex 9: Adapters sequence

Overhang	Sequence (5'→3')
Forward	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
Reverse	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG